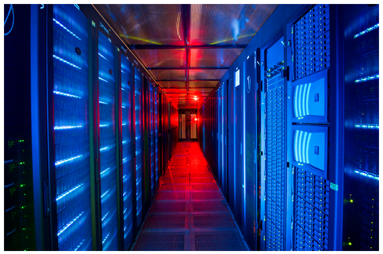
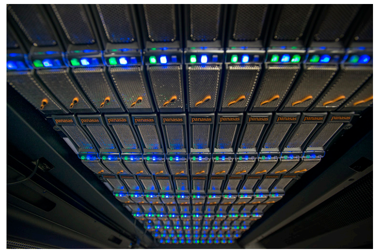
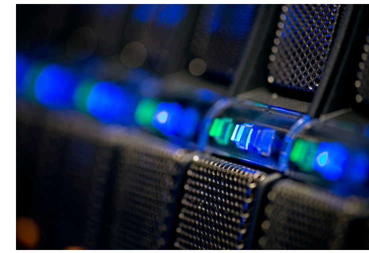
Science & Technology Facilities Council
Rutherford Appleton Laboratory



JASMIN (STFC/Stephen Kill)

JASMIN and the role of Cloud Computing in realising a Big Data facility for the Environmental Sciences

European Conference for Applied Meteorology and Climatology 2017

4–8 September 2017, Dublin, Ireland

Philip Kershaw[+]

Victoria Bennett[+], Jonathan Churchill[*], Bryan Lawrence[^], Matt Pritchard[+], Matt Pryor[+]

+NCAS/NCEO CEDA STFC, *STFC Scientific Computing Dept., ^NCAS / University of Reading

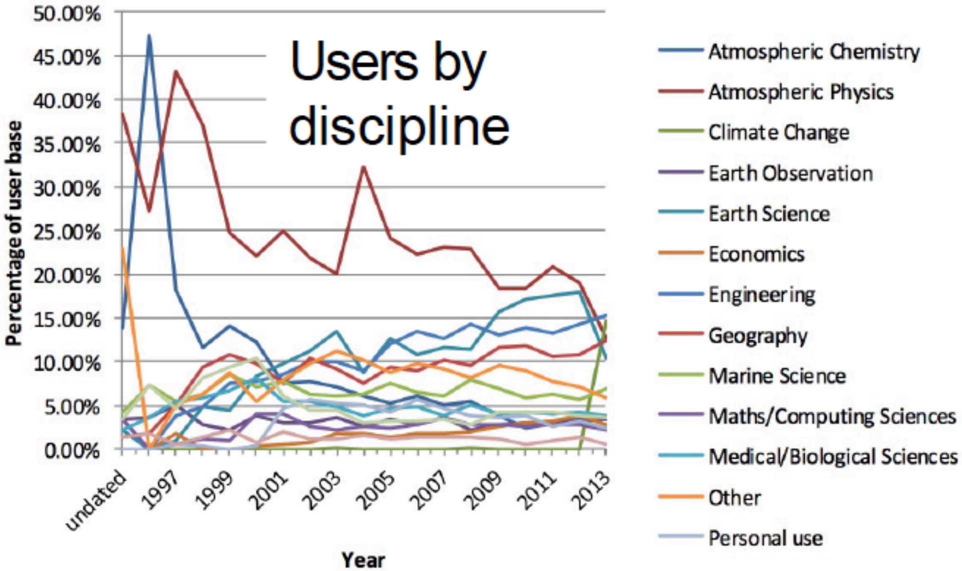
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

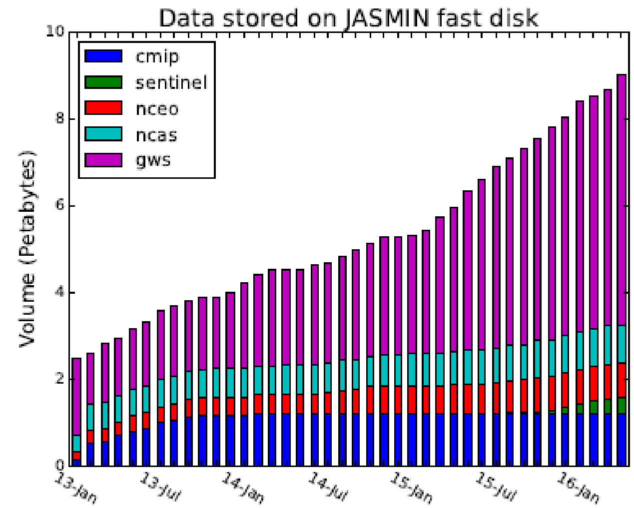
Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council
Rutherford Appleton Laboratory

All about data: CEDA Data Centres



> 30,000 registered users



2015-2016 increasing data storage on JASMIN, in Group Workspaces (GWS) and archive

CEDA *Curated* Data

Data - Type	Data Volume (Petabytes)
Earth Observation	1.5
Atmospheric Science	0.8
Climate Model	1.2
Total	3.0

- ~ 400 datasets
- ~ 150 million files

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

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

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Evolution in models for data distribution and analysis

Big Data driving changes in architecture ➔

Single data centre

- Download to user desktop model
- CEDA (< 2008: pre-ESGF and pre-JASMIN)

Federated data centres

- Multiple organisations
- Supports geographically distributed download to client environment
- **Earth System Grid Federation** (from 2008)

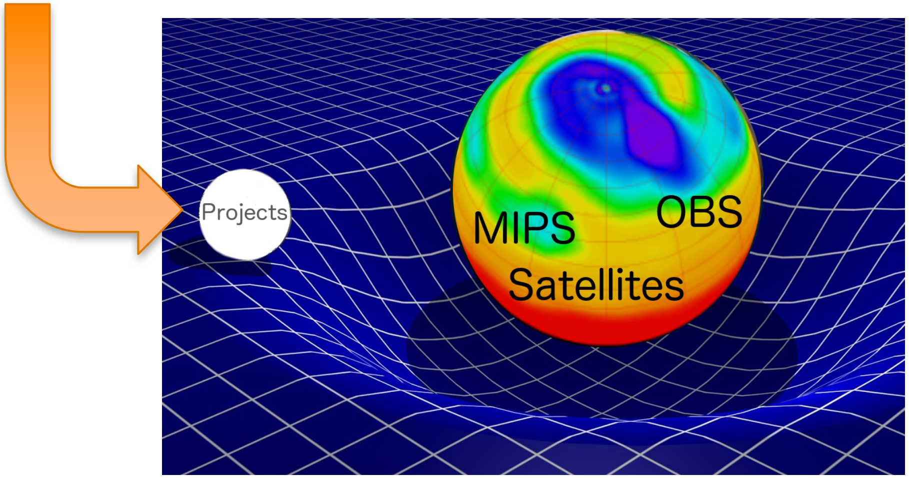
Data analysis facility

- Bring the compute to the data paradigm
- **JASMIN** (from 2012)

JASMIN as a Data Commons

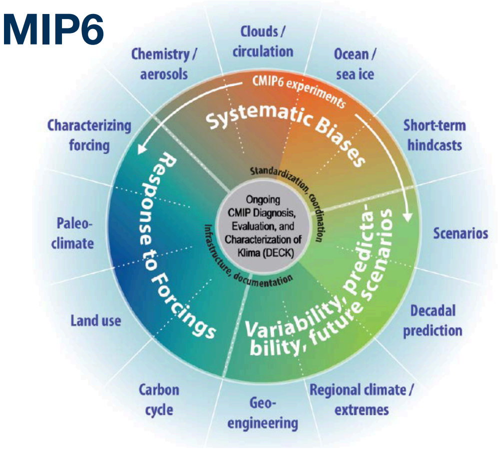
Copernicus
Europe's eyes on Earth

Data gravity associated with *managed data* so that users want to bring their projects to the the JASMIN environment



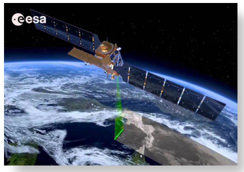
WCRP
World Climate Research Programme

CMIP6



European contribution to HiresMIP alone is expected to exceed 2 PB

Sentinel Data



Sentinel missions data rate: ~6PB/year

Data commons – bring the compute to the data

Make your own VMs, storage and network config

Lotus Batch Compute

Virtualisation

Platform as a Service

Infrastructure as a Service

Software as a Service

e.g. OPeNDAP, WMS, Jupyter Notebook, Web portals

Build and run code to run in parallel on shared environment

CEDA Archive – Data Centres

Group Workspaces - User managed

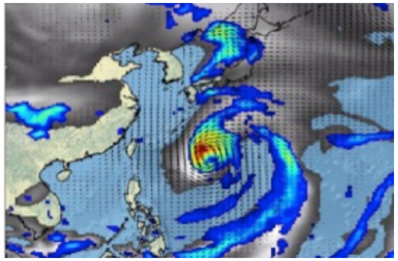
Provision your own shrink-wrapped VMs e.g. database, analysis, web server

Access pre-prepared scientific analysis hosts – preconfigured with apps and libraries for community

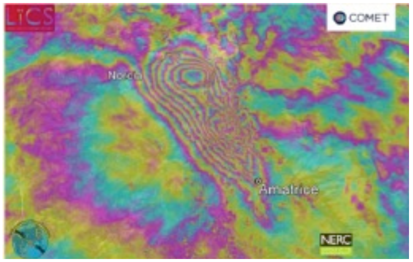
JA... ve Computer

- 19PB Disk storage incl. parallel file system;
- high–speed, non-blocking, low latency networking;
- Batch Compute; Community Cloud; Tape storage;
- Dedicated Data Transfer Zone (DTZ)

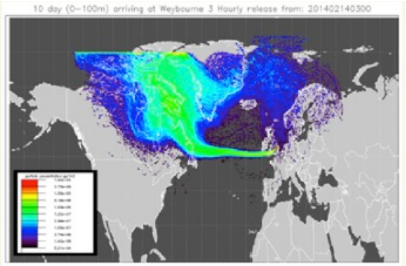
JASMIN usage



High Res Climate Model

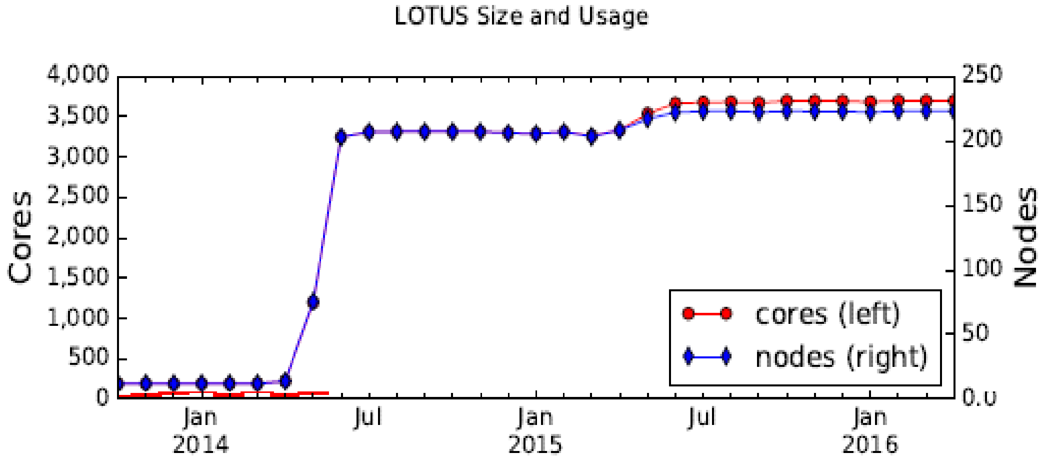


Fault analysis



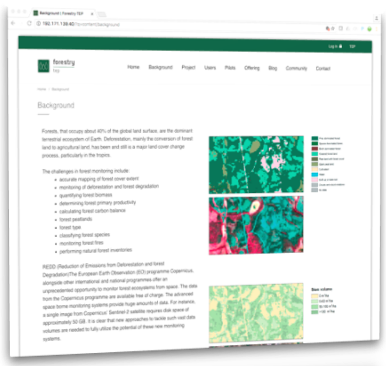
Atmospheric dispersion

- ~150 Science projects on JASMIN
- Lotus, the batch compute environment has had a high-level of utilisation and has been successively expanded over the course of JASMIN's existence.

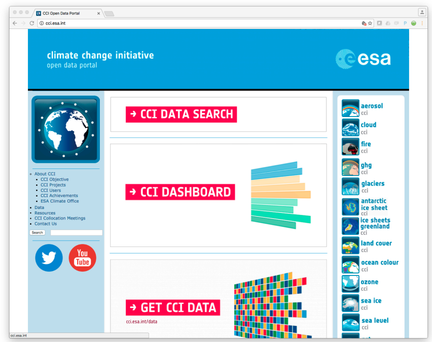


Evolving usage of the Lotus batch cluster: by mid-2016 over 2 million core hours per month!

JASMIN usage: Cloud

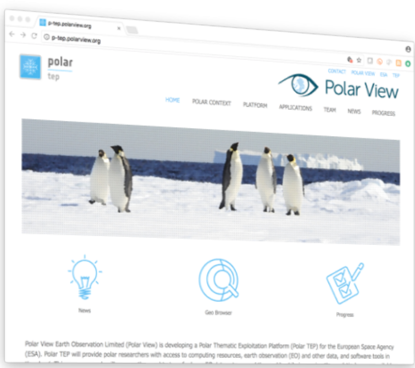


ESA Forestry Thematic Exploitation Platform



ESA Climate Change Initiative Open Data Portal

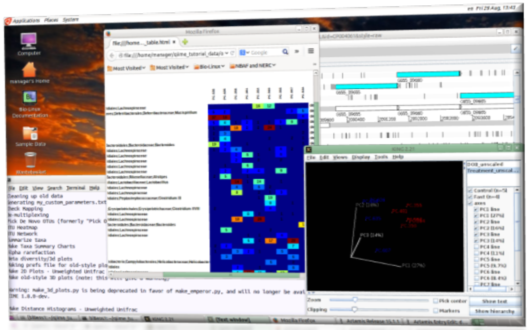
Majic interface to Jules Land-surface model on JASMIN



ESA Polar Thematic Exploitation Platform

Attendees at ESA Summer school, ESRIN used OPTIRAD Jupyter Notebook environment

– Credit ESA



EOS Cloud – Desktop-as-a-Service for Environmental Genomics

Challenges and new developments to address them

Challenges

1) Scaling in terms of
    1) Increasing data volumes
    2) the size of the infrastructure
    3) the numbers of users and
    4) the impact of the above on the overall management and operation of the system.
2) Long-tail science use cases
    1) Effective exploitation of parallelism
    2) More intuitive use – desktop app style experience

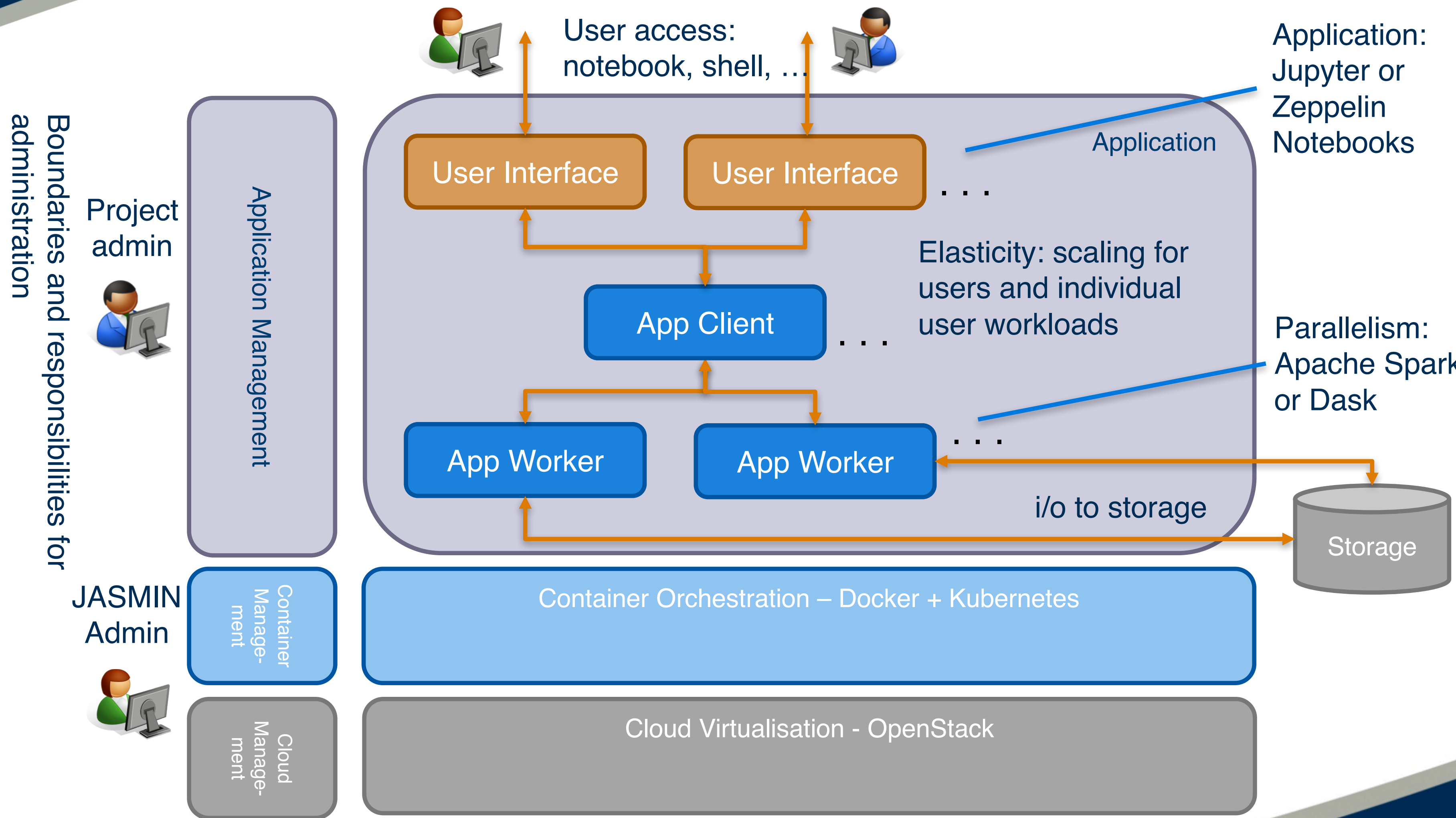
New developments

1) Addressing scaling challenges
    1) Object storage
    2) Container technologies
2) 'Cluster-as-a-Service'
    1) Cluster computing to support
    2) SaaS, Virtual research environments

Cluster-as-a-Service

- Provide cluster to host systems which to some degree abstract parallelism from the user e.g.
  - Apache Spark
  - Dask (Continuum Analytics)
- Provide support for elasticity – cluster resources can be scaled to meet demand and released when unused
- Work is being underpinned by use of container technologies
  - Docker and Kubernetes
  - Provide means to more easily scale and manage services
- Collaborations with
  - MetOffice Informatics Lab (JADE system)
  - NERC DataLab project (Spark with Jupyter/Zeppelin Notebooks)

Cluster-as-a-Service

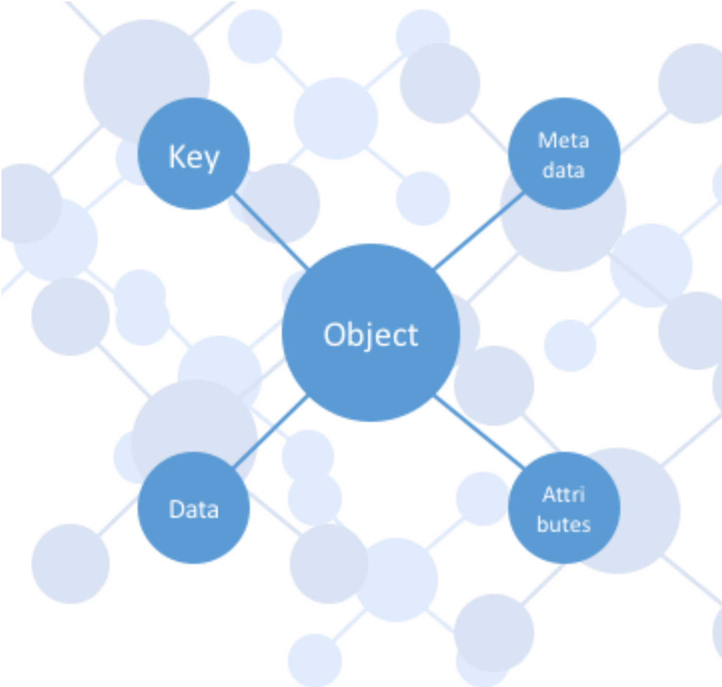



JASMIN (STFC/Stephen Kill)

Data Storage

OBJECT STORAGE

Object Store



- A computer storage architecture in which *Objects* are stored in a flat structure
- Objects are identified by a unique key (a URL)
- Objects are organised into *Buckets*
- Object store can be accessed over a HTTP interface
  - Amazon's S3 HTTP REST API is the most popular
  - Data is uploaded and downloaded using PUT and GET operations respectively
- Data may be split across objects
- Supports two levels of metadata
  - System level metadata
  - Extendable metadata
- Allows searching for data without opening the file and custom searches for user data

Migrate to Object Storage?

POSIX file system

- Fast disk is too expensive at scale
  - Cost of expansion (more PBs)
  - End-of-life for JASMIN1 storage is the end of 2017 (5PB)
- Metadata - minimal
- uid/gid management:
  - bleeding of root privileges beyond scope of a given host to global file system permissions
- Creates unnecessary division in access model – between POSIX and network access protocols such as HTTP

Object Storage

- Cost / TB of object stores is lower
- Metadata
  - Allows searching of files without opening the file
  - As long as the metadata is well constructed
- Independent access model to POSIX uid/gid
- Single common HTTP API (S3) for access – from within infrastructure, from private cloud or external access
- More amenable to cloud architecture: e.g. more easily move apps and data between on-prem and public cloud

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Object Store interfaces

- This represent a fundamental change in access – how do we manage the *interface* to applications and users?

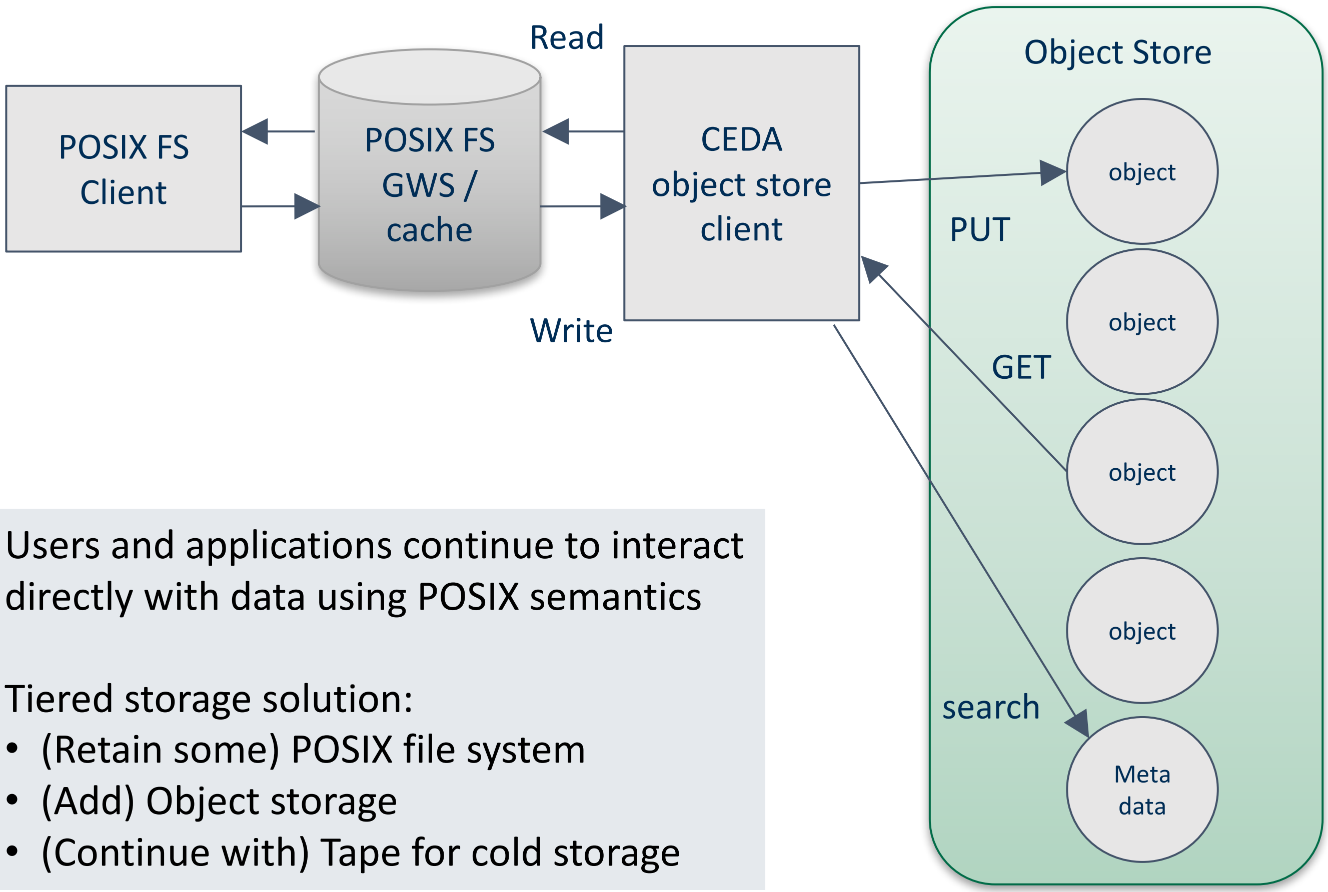
- Two key approaches:

1. S3 interface with POSIX cache
    - shorter timescale (within year)

2. HDF REST API or OPeNDAP over object store:
    - longer term (one year+ even for first datasets)

Science & Technology Facilities Council
Rutherford Appleton Laboratory

1) S3 Interface with POSIX Cache



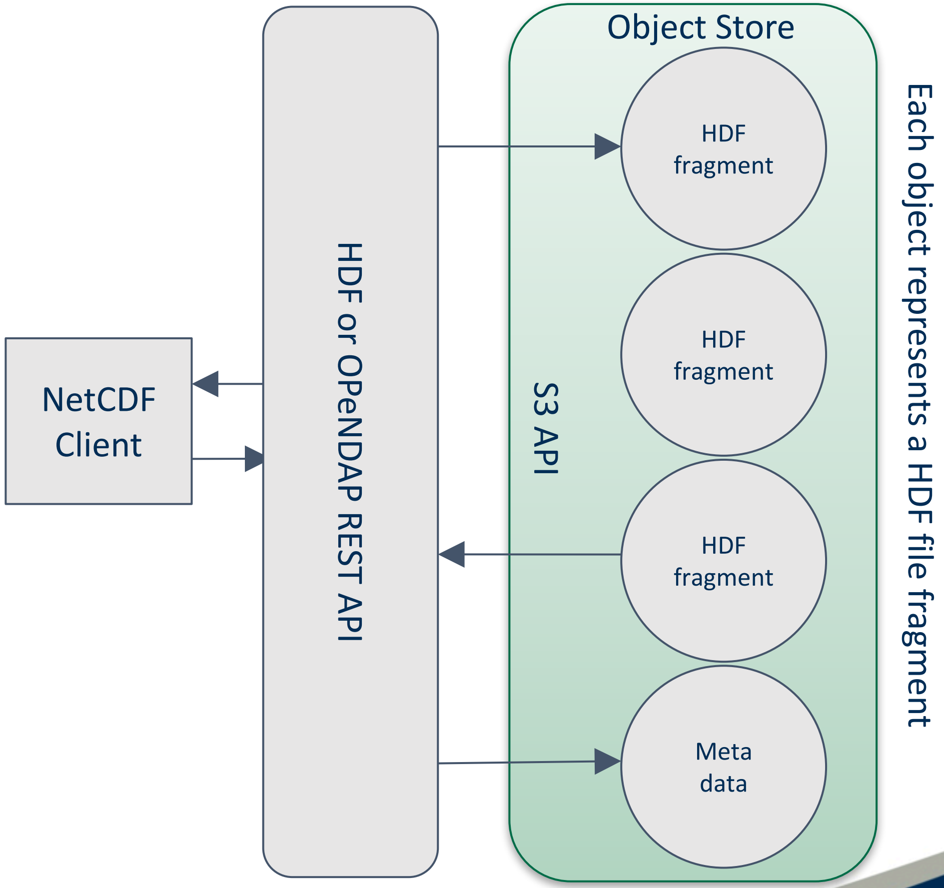
Users and applications continue to interact directly with data using POSIX semantics

Tiered storage solution:

- (Retain some) POSIX file system
- (Add) Object storage
- (Continue with) Tape for cold storage

2) HDF REST API or OPeNDAP over object store

- Data can be split across objects
  - Use HDF data model (underpins netCDF)
  - e.g. netCDF file split into a single timestep / level per object
- Read slices of netCDF data without having to fetch the entire file
- Two possible approaches under development
  - OPeNDAP + CFPython (CEDA in-house)
  - HDF Scalable Data Service (the *HDF Group*)
- Deploy initially for the CEDA curated datasets and then expand to support users' group workspaces
- This methodology may be applied to other file formats with good metadata e.g. Sentinel data



Object Store - Roadmap

- Proof-of-concept with object store vendors just completed
- Further tests being carried out to assess functionality and performance for workloads in our domain
- Rollout object storage in stages:
  - S3 interface with POSIX cache (within year)
  - HDF REST API or OPeNDAP over object store (longer term: one year+ even for first datasets)
- Engage closely with user community to support them with the transition

Summary

- JASMIN: data gravity, a data commons for environmental sciences
- Established system and track record of use by the community
- Challenges with respect to running at scale:
  - Data volumes
  - Numbers of users
- New infrastructure services to address challenges:
  - Cluster-as-a-Service
  - Container-based automation
  - Object store migration
  - Staged rollout to minimise the impact of changes on the user community

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Further Information

- CEDA and JASMIN:
  - http://www.jasmin.ac.uk/
  - http://www.ceda.ac.uk/
- JASMIN paper
  Lawrence, B.N. , V.L. Bennett, J. Churchill, M. Juckes, P. Kershaw, S. Pascoe, S. Pepler, M. Pritchard, and A. Stephens. **Storing and manipulating environmental big data with JASMIN.** *Proceedings of IEEE Big Data 2013, p68-75,* doi:10.1109/BigData.2013.6691556
- ESA Climate Change Initiative Open Data Portal
  - http://cci.esa.int/
- CEDA ESGF node
  - https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/
- ESGF ICMWG (International Climate Network Working Group)
  - http://icnwg.es.net/
- ESNet Science DMZ
  - http://fasterdata.es.net/
- philip.kershaw@stfc.ac.uk, @PhilipJKershaw

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

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