UK academic infrastructure to support (big) environmental science
or
Data Driven Science
Bringing Computation to the Data

Bryan N Lawrence
NCAS Director of Models and Data
Infrastructure, Scope for Today

Shared Infrastructure: Contracts & Service Level Agreements

Formal Collaboration: MoU

Ad-Hoc Collaboration: Trust

Integration

Co-ordination

Management and Governance

Pathways to Impact

Capability

People Development

Computing and Data Skills

Research And Sector Domain Knowledge

Connections

Networks

Security and Authentication

Infrastructure

Data Infrastructure

Hardware And Compute

Software Development

Infrastructure to support (big) environmental science

Bryan Lawrence - Annecy, September 2015
Infrastructure, Scope for Today

UK ACADEMIC Focus
Outline

- Institutional Environment
- Key Drivers
- Data Intensive Computing - JASMIN
- Futures
UK Research Councils and NERC Centres

Infrastructure to support (big) environmental science
Bryan Lawrence - Annecy, September 2015
Infrastructure to support (big) environmental science
Bryan Lawrence - Annecy, September 2015
National Centre for Atmospheric Science

Institutions

- National Centre for Atmospheric Science
- Uni of Leeds
- NCAS Directorate
- NCAS Atm Phys Lead
- Uni of York
- NCAS Composition Lead
- STFC Rutherford Lab
- CEDA (NCAS, NCEO)
- Uni of Manchester
- NCAS Observations Lead
- Cranfield Uni - FAAM
- Uni of Reading
- NCAS Climate Lead
- NCAS M&D Lead

Infrastructure to support (big) environmental science

Bryan Lawrence - Annecy, September 2015
Institutions
NCAS and Computation

Not all NCAS sites shown!

Uni of York
NCAS Composition Lead

Uni of Leeds
NCAS Directorate
NCAS Atm Phys Lead

Cranfield Uni - FAAM

Uni of Reading
NCAS Climate Lead
NCAS M&D Lead

EPCC
ARCHER

STFC Rutherford Lab
CEDA (NCAS, NCEO)

Uni of Manchester
NCAS Observations Lead

Met Office

STFC (NERC)

Infrastructure to support (big) environmental science
Bryan Lawrence - Annecy, September 2015
Computation and Networks

Not all NCAS sites shown!

EPCC
ARCHER

SuperJanet Backbone
(10 Gb/s)
+ 2 Gb Lightpath
+ upgrade plans

1 Gb Lightpath
+ upgrade plans

Met Office

United Kingdom

1 Gb Lightpath

STFC (NERC)
Institutions

National Services delivered by EPCC on behalf of EPSRC and NERC.

- NERC has roughly a fifth of the machine for a total annual allocation of 3.2B AU (213 million core hours)
- extra available via a leadership call (NERC ≈ 30 MCH in 2015).

Compute: Archer Cray XC-30
- 118,080 cores.
- 4920 nodes, each with 2 x 12 core Ivy Bridge (2.7 GHz E5-2697v2),
- Standard nodes (4544) have 64 GB, and “High” Memory nodes (376) 128 GB.
- Aries dragonfly Interconnect.
- I don’t care about the Linpack performance!

Storage: Archer and the Research Data Facility (RDF):
- Archer: /home: NetApp, NFS, 200 TB
- Archer: /work: Sonexion, Lustre, 5 PB
- RDF: /nerc RDF connected by dual 40 Gbit links: DDN GPFS 14 PB with additional backup capacity. Long term storage, but not curated.
Three views of NERC usage on ARCHER from six months ending in March 2015:

- Dominated by climate, atmospheric and oceanic science.
- Unified Model will be both NWP and Climate scale jobs.
- NEMO is the ocean.
- Oasis is the coupler, so those are coupled ocean/atm jobs.
- (VASP is mineral physics.)
MONSoon — JWCRP Shared Development Platform

Institutional Environment

Met Office main platform now a Cray XC-40 (recently migrated from IBM Power).

- Academic community have no direct access to MetO main platforms, and historically have not shared the same HPC architecture.
- Also, historically, no shared access to an analysis environment.

Key Drivers

JWCRP has requirement for shared development platform: MONSoon

- 3712 cores
- 116 nodes, each with 2x16 core Haswell (2.3 GHz)
- 128 GB per node with Aries dragonfly interconnect
- 670 TB Lustre

Summary

Infrastructure to support (big) environmental science

Bryan Lawrence - Annecy, September 2015
Initially conceived of as a response to the JWCRP need for a shared analysis platform. Now much, much more than that...
JASMIN Services

CEDA AS (once BADC)
CEDA EO (once NEODC)
CEDA Solar (once UKSSDC)
IPCC DDC
etc

NERC Managed Analysis Computing
(CEMS + Shared Systems for NCAS, MetO, NOC etc)

NERC Cloud Analysis Computing
(EOS Cloud, Env WB etc)

CEDA Archive Services
Data Centres, Curation, DB systems
User management, External Helpdesk

CEDA Compute Services
Compute Cloud:
PaaS (JAP + Generic Science VMs + User Management), IaaS
External Helpdesk

JASMIN Compute and Storage
Lotus + Private Cloud + Tape Store + DMZ for data transfer
Internal Helpdesk
International Context

- MONSoOn T2
- ARCHER T0
- RDF
- Uni T2/3
- MASS
- 10G backbone & lightpaths
- RAL to TVN to JANET
- jasmin
- Web
- CPU
- Cloud
- ESGF Archive
- Elastic Tape
- PRACE T0
- ESA-PAFs
- IS-ENES
- KNMI
- DKRZ
- GEANT
- ES.NET
- PCMDI
- GFDL
- NCAR
- 1G

KEY: CEDA: □  UK: □  Europe: □  Intercontinental: □
International Context

The network view is the easy view!

▶ What are the data policies?
▶ What are the (possible) data residence times?
▶ What agreements are in place?
▶ What can we rely on in this picture? For example, who has to agree to upgrade something (a network link for example)?

▶ How do community science drivers/requirements lead to infrastructure provision.

▶ All out of scope for today!
Where is this going?

- (Potentially) many different remote simulation sources. How long can the data remain at source?
- Interesting problems moving the data to a common location?
- How long can the data reside on disk at the analysis location? What about in the archive?
- How should we best organise the data?
- What are the best ways to organise analysis compute?
- What are the best ways to address analysis interconnect and I/O bandwidth?
Programmes and Models

Earth System Modelling
PI C. Jones (NCAS at the Met Office)

High Resolution Climate Modelling
Joint PIs: P-L. Vidale (NCAS), M. Roberts (Met Office)

Essentially the same physics/dynamics parameters used throughout model hierarchy

UPSCALE 2012-2013
PRIMAVERA including CASIM – Planned 2015-16

Project to assess impact of global explicit convection

GloSea5
UKESM
Lo   Hi

CMIP5 resolution

Joint Weather and Climate Research Programme
A partnership in climate research

Infrastructure to support (big) environmental science
Bryan Lawrence - Annecy, September 2015
The Science

The Propagation of Direct Numerical Simulation

Primarily mathematical representation of a complex system of processes

Image: from J. Laleuile, 2006

More communities want to observe and simulate the world at ever higher resolution!

More complexity!

http://www.bgs.ac.uk/research/environmentalModelling/home.html
Many interacting communities, each with their own software, compute environments etc.

Figure adapted from Moss et al, 2010
ESGF

Earth System Grid Federation

Downstream Users

Bryan Lawrence - Annecy, September 2015
The trend

Individual End Users
- Limited resources (bandwidth, storage, ...)

Organized User Groups
- Organize a local cache of required files
- Most of group don't access ESGF, use cache instead!

Data Centre Service Group
- Provides access to ESGF replica cache
- May also provide access to data near compute resources
- (BADC, DKRZ, IPSL, KNMI, UC)

Trend

Needed: Replacement for „Download and Process at Home“ Approach
Faster Compute


2014: Archer
Faster Compute


2014: Archer

EPCC Advanced Computing Facility, 2014
Faster Compute


2014: Archer

EPCC Advanced Computing Facility, 2014

From 1981, without Moore’s Law

Slide content courtesy of Arthur Trew
More Data

Fig. 2 The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what’s needed, particularly if you’re not a climate scientist.

(BNL: Even if you are?)
Doing things with Data: Sentinel 1

Sentinel 1A: Launched 2014 (1B due 2016)
- Key instrument: Synthetic Aperture Radar
- Data rate (two satellites: raw 1.8 TB/day, archive products ~ 2 PB/year)

COMET: Centre for Observation and Modelling of Earthquakes, Volcanoes, and Tectonics

(Picture credits: ESA, ArianeSpace.com, PPO.labs-Norut-COMET-SEOM Insarap study, ewf.nerc.ac.uk/2014/09/02/new-satellite-maps-out-napa-valley-earthquake/)
 Doing things with Data: Sentinel Data Rates

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Launch Dates</th>
<th>Daily Data Rate</th>
<th>Product Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A, S1B</td>
<td>Apr 2014</td>
<td>1.8 TB/day raw</td>
<td>2 PB/year</td>
</tr>
<tr>
<td>S2A, S2B</td>
<td>Jun 2015</td>
<td>1.6 TB/day raw</td>
<td>2.4 PB/year</td>
</tr>
<tr>
<td>S3A, S3B</td>
<td>Oct 2015</td>
<td>0.6 TB/day raw</td>
<td>2 PB/year (L1,L2,L3)</td>
</tr>
</tbody>
</table>

with more satellites in the pipeline. Too easy to say “petabytes”!
Doing things with Data: Sentinel Data Rates

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Launch Dates</th>
<th>Daily Data Rate</th>
<th>Product Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A, S1B</td>
<td>Apr 2014</td>
<td>1.8 TB/day raw</td>
<td>2 PB/year</td>
</tr>
<tr>
<td>S2A, S2B</td>
<td>Jun 2015</td>
<td>1.6 TB/day raw</td>
<td>2.4 PB/year</td>
</tr>
<tr>
<td>S3A, S3B</td>
<td>Oct 2015</td>
<td>0.6 TB/day raw</td>
<td>2 PB/year (L1,L2,L3)</td>
</tr>
</tbody>
</table>

with more satellites in the pipeline. Too easy to say “petabytes”!

- Traditional approach: write data to tapestore, users retrieve scenes from a catalogue!
- Modern “big data” approach: users want to do “whole mission” reprocessing!
  - e.g. QA4ECV (J-P Muller): bought 800 TB of disk in the JASMIN system, now running whole mission reprocessing 100x faster than their in-house cluster. Days to test new science instead of months. Massive improvement in scientific throughput!
“Without substantial research effort into new methods of storage, data dissemination, data semantics, and visualization, all aimed at bringing analysis and computation to the data, rather than trying to download the data and perform analysis locally, it is likely that the data might become frustratingly inaccessible to users”

A National Strategy for Advancing Climate Modeling, 2012

Semantic Analysis: “substantial research effort” “new methods” “computation to data” “rather than trying to download” “frustratingly inaccessible” (to whom?)
Sharing

Science across scales

Lots of interacting communities

Lots of infrastructure

New sorts of infrastructure

Can we share infrastructure?
So we have built a Data Intensive Computing System: JASMIN

- 16 PB Fast Storage
  (Panasas, many Tbit/s bandwidth)
- 1 PB Bulk Storage
- Elastic Tape
- 4000 cores: half deployed as hypervisors, half as the “Lotus” batch cluster.
- Some high memory nodes, a range, bottom heavy.
Virtual Organisations

Platform as a Service → Infrastructure as a Service

Example: NCAS will run a semi-managed virtual organisation (with multiple group work spaces), but large groups within NCAS can themselves also run virtual organisations.
High performance, curation + facilitation

Objective is to provide an environment with high performance access to curated data archive and a high performance data analysis environment!

- **CEDA**
  - Virtual Machines
  - Archive Storage

- **Your Managed VO**
  - Virtual Machines
  - VO Fast Storage

- **Your UnManaged VO**
  - Virtual Machines
  - VO Storage

**Legend:**
- **RW:** Data mounted Read/Write
- **RO:** Data mounted ReadOnly
- **SaaS:** Data access via Software as a Service (OpenDAP etc)
Currently $\mathcal{O}(100)$ “Group Work Spaces” in the managed cloud serving $\mathcal{O}(100)$ “virtual organisations” and $\mathcal{O}(500)$ users (there is some overlap). Unmanaged cloud is currently in testing with a few brave souls.
Integrated Cloud Provisioning 2

- IPython Notebook VM could access cluster through Python API
- SSH via public IP
- CloudBioLinux Desktop

JASMIN Cloud Management Interfaces

Managed Cloud - PaaS, SaaS
- Science Analysis VM
- Storage
- Project1-org
- Direct access to batch processing cluster
- Lotus Batch Compute

Managed Cloud - PaaS, SaaS
- Login VM
- Science Analysis VM
- Project2-org
- Direct File System Access
- Data Centre Archive

Unmanaged Cloud
- Compute Cluster VM
- CloudBioLinux Fat Node
- File Server VM
- Storage
- eos-cloud-org

External Network inside JASMIN

IaaS, PaaS, SaaS


High Performance Network
It’s easy to access and exploit the managed archive from user environments in the managed cloud!
JASMIN Hosted Processing and Archive Access

It's easy to access and exploit the managed archive from user environments in the managed cloud!
Accessing data: The Status Quo

Users in the managed cloud have file system access. Users in the JASMIN private cloud could get access directly to data in the managed cloud, but that’s not secure. So, they need to access the data through software interfaces (software as a service, SaaS), or copy (aka, download) the data locally:

This requires managed services in the managed cloud and requires data duplication outside for file transfer. OPeNDAP may be hard to make scalable and performant. What if it wasn’t a POSIX file system?
We are investigating, with the HDF group, whether we can build a performant (compared with PanFS) HDF interface for reading data at scale (we may or may not want different solutions for the archive and the GWS).

If successful, we could replace the necessity for running OPeNDAP servers, and we could exploit (cheaper, denser) object storage via the regular netcdf4 libraries.

We’re currently investigating CEPH.

This work will complement our plans under ESIWACE!
A Hybrid Cloud Future?

- It’s clear that we cannot provide compute at comparable scale to the public cloud.
- It’s also clear that we need to simplify provisioning of cloud resources for our tenants.
- Solution: Develop our own cloud federation portal: “cloudhands”! (it is clear that we are far from a “industry standard API”).
- In the long run we want to see workflow that straddles the hybrid cloud, exploiting “academic” data intensive computing (itself downstream from sensors and HPC) and “public” generic computing where the academic provision is not adequate.
Final Remarks

▶ The UK academic computing environment is getting more complicated as we

1. move away from the “one remote HPC download to my departmental compute” mode, and
2. face both much more interdisciplinarity, alongside
3. much more heterogeneity in the hardware and software of our workflow
Final Remarks

- The UK academic computing environment is getting more complicated as we
  1. move away from the “one remote HPC download to my departmental compute” mode, and
  2. face both much more interdisciplinarity, alongside
  3. much more heterogeneity in the hardware and software of our workflow

- After a some decades, and with the advent of the cloud, we are much closer to realising the “bringing compute to the data” vision that “The Grid” espoused but could never quite deliver.
The UK academic computing environment is getting more complicated as we
1. move away from the “one remote HPC download to my departmental compute” mode, and
2. face both much more interdisciplinarity, alongside
3. much more heterogeneity in the hardware and software of our workflow

After a some decades, and with the advent of the cloud, we are much closer to realising the “bringing compute to the data” vision that “The Grid” espoused but could never quite deliver.

We are starting to recognise the necessity to work much closer with those building both key elements of our software stack (e.g. The HDF Group) and of our hardware stack (especially storage vendors)
Final Remarks

- The UK academic computing environment is getting more complicated as we
  1. move away from the “one remote HPC download to my departmental compute” mode, and
  2. face both much more interdisciplinarity, alongside
  3. much more heterogeneity in the hardware and software of our workflow

- After a some decades, and with the advent of the cloud, we are much closer to realising the “bringing compute to the data” vision that “The Grid” espoused but could never quite deliver.

- We are starting to recognise the necessity to work much closer with those building both key elements of our software stack (e.g. The HDF Group) and of our hardware stack (especially storage vendors)

- Container technology (Docker, Mesos and friends) will shake up our cosy plans for the future!