The NCI High Performance Computing and High Performance Data Platform to Support the Analysis of Petascale Environmental Data Collections

Ben Evans, Lesley Wyborn, Tim Pugh, Chris Allen, Joseph Antony, Kashif Gohar, David Porter, Jon Smillie, Claire Trenham, Jingbo Wang, et al.

The NCI High Performance Computing and High Performance Data Platform to Support the Analysis of Petascale Environmental Data Collections

Ben Evans[1], Lesley Wyborn[1], Tim Pugh[2], Chris Allen[1], Joseph Antony[1], Kashif Gohar[1], David Porter[1], Jon Smillie[1], Claire Trenham[1], Jingbo Wang[1], Alex Ip[3], Gavin Bell[4]

[1] National Computational Infrastructure (NCI), Australian National University, Canberra, Australia
[2] Bureau of Meteorology, Melbourne, Australia
[3] Geoscience Australia, Canberra, Australia
[4] The 6th Column Project, Berlin, Germany
{Ben.Evans, Lesley.Wyborn, Chris.Allen, Joseph.Antony, Kashif Gohar, David.Porter, Jon.Smillie, Claire.Trenham, Jingbo.Wang}@anu.edu.au
T.Pugh@bom.gov.au
Alex.Ip@ga.gov.au
gavin@6thcolumn.org

Abstract. The National Computational Infrastructure (NCI) at the Australian National University (ANU) has co-located a priority set of over 10 PetaBytes (PBytes) of national data collections within a HPC research facility. The facility provides an integrated high-performance computational and storage platform, or a High Performance Data (HPD) platform, to serve and analyse the massive amounts of data across the spectrum of environmental collections – in particular from the climate, environmental and geoscientific domains. The data is managed in concert with the government agencies, major academic research communities and collaborating overseas organisations. By co-locating the vast data collections with high performance computing environments and harmonising these large valuable data assets, new opportunities have arisen for Data-Intensive interdisciplinary science at scales and resolutions not hitherto possible.

Keywords: high performance computing · high performance data · cloud Computing · data-intensive science · scalable data services · data cube · virtual laboratories

1 Introduction

The National Computational Infrastructure (NCI) at the Australian National University (ANU) has organised a priority set of large volume national environmental data assets on a High Performance Data (HPD) Node within a High Performance Compu-

ting (HPC) facility, as a special node under the Australian Government's National Collaborative Research Infrastructure Strategy [1] Research Data Storage Infrastructure (RDSI) program [2]. The colocation of these large volume collections with a high performance and flexible computational infrastructure is part of an emergent area of the Data-Intensive Science, at times referred to as the Fourth Paradigm of Science [3].

NCI operates as a formal partnership between the ANU and the three major Australian National Scientific Agencies: the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the Bureau of Meteorology (BoM) and Geoscience Australia (GA) who are also the custodians of many of the large volume national scientific data records. The data from these national agencies and collaborating overseas organisations are either replicated to or produced at NCI, and in many cases processed to higher-level data products. Model data from computational workflows at NCI are also captured and released as modelling products. NCI then manages both data services and computational environments, known as Virtual Laboratories, to use that data effectively and efficiently.

This paper examines the data collections, the underlying High Performance Data (HPD) infrastructure, the management of trusted environments, and opportunities and challenges that this new research platform at NCI offers.

2 Data Collections

There are 31 (and growing) data collections in the initial ingestion at NCI [4] requiring over 10 Petabytes (PBytes) in storage volume (Table 1). They are currently categorised into six major fields all related to the environmental sciences:

1) earth system sciences, climate and weather model data assets and products;
2) earth and marine observations and products;
3) geosciences;
4) terrestrial ecosystem;
5) water management and hydrology; and
6) astronomy, social science and biosciences.

These data collections are predicted to grow exponentially over time. For example, the datasets for the international Coupled Model Inter-comparison Project (CMIP5) [5] experiment completed in 2013 are well over 3 PBytes; the data for the next CMIP6 experiment will be an order of magnitude larger again by 2020 due to higher resolution spatial grids. Current satellite remote sensing data collections, along with their higher products derived from large time-series of Australian Landsat collection, Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High Resolution Radiometer (AVHRR) and Visible Infrared Imaging Radiometer Suite (VIIRS) are in the order of 2 PBytes. The next generation of satellites such as Japanese Himawari-8/9 mission [6] and the European Space Agency's Copernicus programme [7] are about to come online, with higher spatial resolution sensors and data volumes that in time will dwarf the earlier satellite data collections.

To facilitate effective and efficient use of large volume data collections and to support their use in high performance computing environments, attention must be placed

Table 1. Data Collections hosted at the NCI RDSI node as of 27 November, 2014. The fields are 1) earth system sciences, climate and weather model data assets and products, 2) earth and marine observations and products, 3) geosciences, 4) terrestrial ecosystem, 5) water management and hydrology, and 6) astronomy, social science and biosciences.

Field	Collection Name	TBytes
1	Ocean General Circulation Model for the Earth Simulator	27
1	Year Of Tropical Convection (YOTC) Re-analysis	17
1	Community Atmosphere Biosphere Land Exchange (CABLE) Datasets	23
1	Coordinated Regional Climate Downscaling Experiment (CORDEX)	52
1	Coupled Model Inter-Comparison Project (CMIP5)	3077
1	Atmospheric Reanalysis Products	104
1	Australian Community Climate and Earth-System Simulator (ACCESS)	2350
1	Seasonal Climate Prediction	546
2	Australian Bathymetry and Elevation reference data	105
2	Australian Marine Video and Imagery Collection	6
2	Global Navigation Satellite System (GNSS) (Geodesy)	5
2	Digitised Australian Aerial Survey Photography	72
2	Earth Observation (Satellite: Landsat, etc)	1587
2	Satellite Imagery (NOAA/AVHRR, MODIS, VIIRS, AusCover)	152
2	Satellite Soil Moisture Products	5
2	Synthetic Aperture Radar	27
2	Remote and In-Situ Observations Products for Earth System Modelling	341
2	Ocean-Marine Collections	399
3	Australian 3D Geological Models and supporting data	3
3	Australian Geophysical Data Collection	307
3	Australian Natural Hazards Archive	26
3	National CT-Lab Tomographic Collection	190
4	ecosystem Modelling And Scaling faciliTy (eMAST)	84
4	Phenology Monitoring (Near Surface Remote Sensing)	11
5	eMAST Data Assimilation	102
5	Key Water Assets	41
5	Models of Land and Water Dynamics from Space	20
6	Skymapper (Astronomy)	207
6	Australian Data Archive (Social Sciences)	4
6	BioPlatforms Australia (BPA) Melanoma Dataset (Biosciences)	120
6	Plant Phenomics (Biosciences)	102

on the whole data workflow from creation to publication, including data management plans, provenance capture, and unique identification of the data through Digital Object Identifiers (DOIs) and other forms of data discovery and access. As there is also a move towards enabling HPC analytics directly on the data content, increasingly attributes will need to be self-describing and conform to international standards for vocabularies and ontologies.

Because of the scale of the data collections at NCI, some aspects of data management are now active areas of ongoing research and development collaboration. One example is provenance systems, where specialized databases will be created to allow use-dependent provenance graphs to be generated on-the-fly.

3   High Performance Data (HPD)

With both the increased data volumes, and the increased computing power required to effectively process and analyse them, the coupled computational environments are increasingly too large or complicated to be architected separately: they must be co-designed at inception. This has given rise to High Performance Data (HPD), as a complementary approach to the traditional High Performance Computing (HPC).

We define High Performance Data (HPD) as data that is carefully prepared, standardised and structured so that it can be used in Data-Intensive Science on HPC.

To achieve this, the use of the data needs to be considered as part of the production and management of the data, rather than an after-thought for data curation.

Many of the data sets that have been assembled on the NCI are ones that were considered too large to process and access by conventional means; bandwidth limits the capacity to move them, data transfers are too slow; and latencies to access them are too high. Even if the data sets can be moved, few can afford to store them locally, the energy costs are substantial, and provenance and versioning of the large data collections at multiple sites are too complex to manage. We predict the increasing scale of data, its associated processing, and complexity of software environments will be too difficult or affordable to host beyond a few trusted international centres.

Properly structured, the HPD collections provide new opportunities to integrate and process large volumes of data at resolutions and at scales never before possible. Once created, HPD enables in-situ processing of the data and requires users changing their processing to use trusted online applications and services on the data in-situ.

4   High Performance Data Infrastructure

4.1   Design Considerations

The infrastructure at NCI has been designed to support Data-Intensive Science within HPC environments. The new infrastructure was designed to reduce the time taken to process the data and required a balance between the available processing power and the ability to access data.

More specifically the infrastructure to support HPD at NCI was designed and programmed, and is operated in a way that enables multiple users to rapidly invoke dynamic high performance computation on these large volume data collections from a variety of use-cases. Processing can take place in a variety of modes: batch-style deep processing; interactive environments for preparing model and analysis computation; web-time data processing including the Open Geospatial Consortium (OGC) Web Processing Service (WPS) [8], Virtual Desktop Interfaces (VDIs) for fully assembled software environments; and various server-side visualisation capabilities.

4.2   Computer Platforms for HPD

To ensure that HPD techniques can scale out for the demands of Data-Intensive Science, all parts of the data and services need to be considered for their ability to scale out and for high bandwidth and low latency access to data. This includes batch-mode HPC platforms, cloud infrastructure, and high performance filesystems.

NCI provides a 1.2 PetaFLOP supercomputer (Raijin) with over 57,000 processing cores and an approximately 10 PBytes Lustre filesystem. The system includes Fujitsu compute nodes consisting of Intel SandyBridge processors, a QDR InfiniBand network, and a Lustre filesystem for a cluster-wide namespace and a PBSPro batch system.

For its private cloud, (Tenjin), NCI selected OpenStack [9] as its underlying cloud technology on a HPC-class 3000 processor core system using the same generation of technology as the Raijin supercomputer, including Intel's SandyBridge processors and a Mellanox QDR InfiniBand interconnect. In addition, each node of Tenjin has local Solid State Drives (SSDs) to allow high IOPs workloads. When using virtual machines, the performance of the hypervisor is critical for the performance of both computational and IO systems. Over time, the performance of these hypervisors has improved. Future work on OpenStack will see bare-metal nodes used as part of the node structure where virtualisation is the limiting factor for operations.

Both the supercomputer and cloud can mount several facility wide Lustre [10] file systems, which maintain the data collections and other persistent reference data. The Lustre systems use a mixture of NetApp and DDN storage arrays, which are managed by Object Storage Servers (OSS) within scalable units. There are currently over 50 OSSes serving approximately 20 PBytes of filesystem space.

To provide operational robustness, each layer of the infrastructure has its own management domain. The connection to both the supercomputer and the cloud are managed with a scalable set of data serving bridging infrastructure using Lustre LNET and NFS that manages both the performance and security to each of the systems over the InfiniBand network fabric.

4.3   Software Stack Management

The variety of both environments and services for Data-Intensive HPD environmental science has required a deeper consideration of the software environment including management and deployment practices and the supporting software infrastructure.

This involves the two key concepts of the Development-to-Operations (DevOps) process which includes firstly testing and integration, and secondly re-usable trusted software stacks.

To manage the software systems, NCI has developed a system of Puppet Bundles based on Puppet [11] and Git [12]. The NCI Puppet Bundles are run in a masterless mode, where the coordination of nodes and deployment is achieved through standard Git management. All operational stacks are maintained to a high degree of traceability and reproducibility through this system.

All shareable services of the NCI environment have been written to make use of the underlying Puppet core framework. This provides easy integration, simplified sharing and re-use, and the ability to more easily scale-out the infrastructure. Other technologies (e.g., Docker [13]) can be deployed in a layered ecosystem of software deployments.

4.4 Scalable Data Access

Modern Data-Intensive environments require much more than just discovery and access to filesystems. To find data, a research problem requires the ability to programmatically search for attributes and extract data using defined services like the Data Access Protocol (DAP), and other geospatial services such as those defined by the OGC including Web Map Service (WMS) [14], Web Coverage Service (WCS) [15] and Web Feature Service (WFS) [16].

NCI have developed a scalable catalogue framework using a combination of GeoNetwork [17] and ElasticSearch [18] to record the metadata, and allows for rapid search across key attributes and respond with the relevant data service URI or filesystem location for access.

All data is registered in the catalogue framework including the services that are meaningful to the data, e.g., THREDDS [19], Hyrax [20], and GeoServer [21]. To find the data, programs have begun to standardize on JSON [22] objects for interacting with the search infrastructure.

4.5 Datacube Technologies

For analysis in high performance environments, those collections that comprise multiple data sets and data series, even if they conform to agreed international standards, need to be aggregated into seamless data arrays. The datasets being assembled are often of different spatial and temporal resolution. To analyse these data as coherent units, the data will increasingly need to be stored, aligned and indexed using datacube methodologies.

A number of collaborations have now been established to address this issue. For example, GA, CSIRO and NCI are collaborating on a project called the Australian Geoscience Data Cube (AGDC) [23]. For this project, the data is prepared to allow rapid analysis of the data in several dimensions (e.g., spectral, spatial and temporal) by converting to consistent grids, each with a fixed and consistent geospatial footprint, forming a space-time 'data cube'. In the development phase of the AGDC,

some 636,000 Landsat product scenes were converted to 4 million spatially-regular, time-stamped, band-aggregated tiles. This consistent gridding minimised the geospatial processing necessary to locate individual pixels of data across multiple scenes (or files) of Earth observation data. Tile-based analyses could then be distributed across thousands of compute nodes for parallel computation underpinned by the high-performance Lustre filesystems of the NCI facility. As a result of this successful development phase, the AGDC project is now being expanded to incorporate other data Earth Observation imagery including MODIS data and elevation data.

5   Interdisciplinary and Collaborative Virtual Laboratories

Virtual Laboratories are a set of integrated software stacks that provides workflow software, specialised tools, web services, and visualisation services. The nature of the Virtual Laboratories is that they consist of a mixture of software stacks that are sourced from independent, trusted software and data repositories and are integrated as an interoperable stack.

Common software stack components may also be reused, and re-integrated by the individual Virtual Laboratory that requires them. This is a key way where software (and ultimately data) can be reused and/or repurposed for multiple Virtual Laboratories and will ultimately assist in achieving interdisciplinary outcomes.

The development of these trusted software stacks usually involves large-scale engineering efforts, often entailing global collaborations. Some well-known examples used at NCI include the Earth Systems Grid Federation (ESGF) [24] (of which NCI is a core member), and model management environments such as Cylc [25] and Rose [26] and visualisation software such as UV-CDAT [27] and the Oceans Data Interoperability Platform (ODIP) [28].

The release and integration of the software environments at NCI are managed using our Puppet Bundles approach, which ensures that software is both upgradable and maintainable. By carefully managing the software stack to track the upstream software repository, the trustworthiness of individual Virtual Laboratory stacks can be maintained.

6   Opportunities and Challenges

The new infrastructure developed at the NCI has enabled:

- A dramatic improvement in the scale, resolution and reach of Australian environmental research;
- High performance access to nationally significant research collections using new Data-Intensive capabilities to support cutting-edge research methods; and
- The realisation of synergies with related international research infrastructure programs.

Further, by co-locating major diverse environmental data collections with HPC infrastructures an interdisciplinary approach to solving key global Environmental Science issues such as impacts of climate change; sustainably exploiting scarce water, mineral and energy resources; and protecting our communities through better prediction of the behaviour of natural hazards is now feasible.

However, these opportunities cannot be easily realised unless some key remaining challenges are addressed, including common licensing framework, authentication and authorisation, semantic enablement of data using controlled vocabularies, and provenance tracking of workflows. There is also a growing recognition of the need for the development of new specialised skills, in particular in data science. There is also a need to deploy these skills in multi-faceted interdisciplinary teams.

7 Conclusions

We have reached the following conclusions about the future directions of computational and Data-Intensive methods in environmental and earth systems science through the course of this work:

1. The need to access the large and exponentially increasing data volumes in Environmental Science is now acute; traditional data access and HPC technologies no longer scale.
2. We define High Performance Data (HPD) as data that is carefully prepared, standardised and structured so that it can be used in Data-Intensive Science on HPC.
3. HPC can now be combined with HPD to support the more complex workflows of Data-Intensive Science.
4. The coupled HPC-HPD environments and the data are increasingly too large or complicated to be architected separately: they must be co-designed at inception.
5. HPC and HPD are now integral to Earth and Environmental Research. Combined they offer new opportunities to collaboratively address problems such as climate change; sustainable exploitation of our water, mineral and energy resources; and better prediction of natural hazards.

Acknowledgements

References

1. Department of Education: National Collaboration Research Infrastructure Strategy, retrieve from https://education.gov.au/national-collaborative-research-infrastructure-strategy-ncris on 27th November 2014.
2. Department of Education: Information Sheet Research Data Storage Initiative,(RDSI) http://docs.education.gov.au/documents/information-sheet-research-data-storage-initiative-rdsi, *Australian Government*, modified 24 September 2014.
3. Hey, T., Tansley, S. and Tolle, K.: Jim Grey on eScience: a transformed scientific method. In: Hey, T., Tansley, S., and Tolle, K. (eds) The Fourth Paradigm: Data-Intensive Science Discovery, pp xvii-xxxi, Microsoft Corporation, USA, 2009.
4. National Computational Infrastructure: National Collections, retrieved from http://nci.org.au/data-collections/data-collections/ and the GeoNetwork catalogue http://geonetwork.nci.org.au on 27th November 2014.
5. Taylor, K.E., R.J. Stouffer, G.A. Meehl: An Overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., 93, 485-498, doi:10.1175/BAMS-D-11-00094.1, 2012.
6. Japan Meteorological Agency: Himarwari-8/9 of the Meteorological Satellite Center (MSC) of the JMA, retrieved from http://www.data.jma.go.jp/mscweb/en/himawari89/index.html on 27th November, 2014.
7. European Space Agency: The Copernicus programme, retrieved from http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview4 on 27th November, 2014.
8. Open Geospatial Consortium Web Processing Service, OGC, retrieved from http://www.opengeospatial.org/standards/wps on 27th November, 2014.
9. OpenStack: retrieved from http://www.openstack.org/ on 27th November, 2014.
10. Lustre file system, retrieved from http://wiki.lustre.org/index.php/Main_Page on 27th November, 2014.
11. Puppet Labs Inc: retrieved from http://puppetlabs.com/ on 27th November, 2014.
12. Git: retrieved from http://git-scm.com/ on 27th November, 2014.
13. Docker: retrieved from https://www.docker.com/ on 27th November, 2014.
14. Open Geospatial Consortium: Web Map Service, retrieved from http://www.opengeospatial.org/standards/wms on 27th November, 2014.
15. Open Geospatial Consortium: Web Coverage Service, OGC, retrieved from http://www.opengeospatial.org/standards/wcs on 27th November, 2014.
16. Open Geospatial Consortium: Web Feature Service, OGC, retrieved from http://www.opengeospatial.org/standards/wfs on 27th November, 2014.
17. Open Source Geospatial Foundation: GeoNetwork, retrieved from http://geonetwork-opensource.org/, modified 31st January 2014.
18. ElasticSearch, retrieved from http://www.elasticsearch.org/ on 27th November, 2014.
19. THREDDS, retrieved from http://www.unidata.ucar.edu/software/thredds/current/tds/ on 27th November, 2014.
20. OpenDAP Inc: Hyrax, retrieved from http://docs.opendap.org/index.php/Hyrax on 27th November, 2014.
21. Open Source Geospatial Foundation: GeoServer, retrieved from http://geoserver.org/ on 27th November, 2014.
22. JSON, retrieved from http://www.json.org/ on 27th November, 2014.
23. Australian Geoscience Data Cube (AGDC), retrieved from https://github.com/GeoscienceAustralia/agdc/wiki on 27th November 2014.

24. Earth Systems Grid Federation: About the Earth System Grid, retrieved from https://www.earthsystemgrid.org/about/overview.htm on 27th November, 2014.
25. Oliver, H: The Cylc Suite Engine, retrieved from http://cylc.github.io/cylc/html/single/cug-html.html on 27th November, 2014.
26. UK Met Office: Rose, retrieved from https://github.com/metomi/rose/ and http://www.metoffice.gov.uk/research/collaboration/rose on 27th November, 2014.
27. Williams, D.N. et al: UV-CDAT, doi:10.5281/zenodo.12251, 2014, See also http://uvcdat.llnl.gov/.
28. Oceans Data Interoperability Platform (ODIP): retrieved from http://www.odip.org/ on 27th November, 2014.