The Australian Data & Computing Landscape

Computing centres and institutional capacity

Simon O'Toole
AAO Macquarie

The SkyMapper Telescope at Siding Spring Observatory – ANU

The Australian landscape

- Pawsey Supercomputing Centre
- NCI
- OzStar and beyond – Swinburne University
- University specific facilities
- Data Central

Pawsey Supercomputing Centre – Storage

Acacia

- Onsite object storage
- 60PB usable space
- Ceph - S3 compatible

Banksia

- Offline tape store
- 74PB available storage
- 5.7 PB disk cache

Pawsey Supercomputing Centre – Compute

Setonix

- 43 petaFlops
- 463 TB RAM
- 1600+ CPU nodes
- 217,088 CPU Cores
- 168,960 GPU cores
- 3.3PB lustre
- SLURM job control
- Used for general scientific computing as well as astronomy

Pawsey – Cloud services

- Nimbus cloud system
- Can run:
  - Virtual machines
  - Kubernetes clusters

- Linked to some data storage (Banksia/Acacia) but has its own dedicated storage

Pawsey Supercomputing Centre

- Who uses Pawsey facilities?
  - ASKAP and MWA telescopes (incl AusSRC)
  - Non-astronomy projects
  - Merit allocation process for CPU hours
  - (Eventually) SKA-Low, although new processing power required

NCI – Compute

GADI

- 74,880 CPU cores in 1,440 52-core 4th Gen Intel Xeon Scalable processors (code-named Sapphire Rapids)
- 720 compute nodes, each with two CPUs per node
- 369 Terabytes of memory
- NVIDIA 200 Gigabit/second Infiniband HDR interconnect

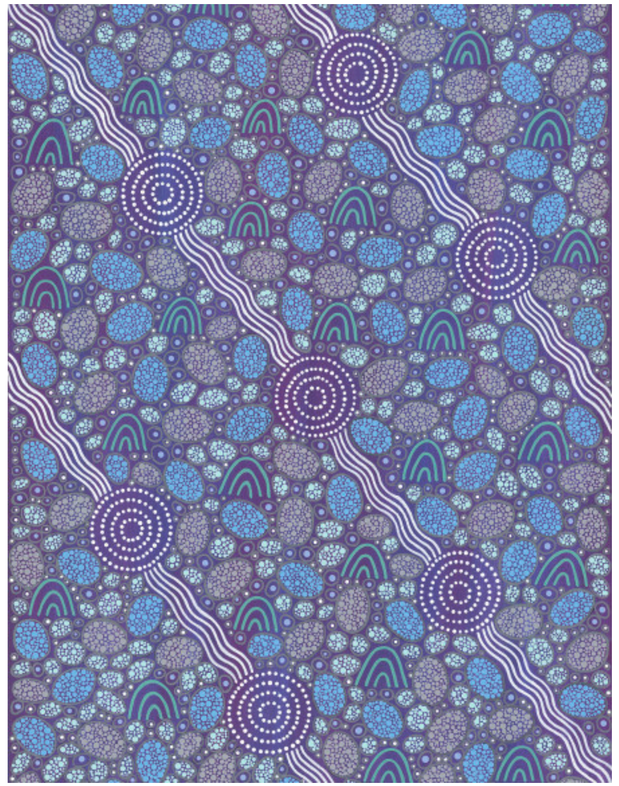
NCI – Storage

Filesystem	Capacity (Petabytes)	Purpose	Availability	Total Aggregate Performance	Maximum Performance
/scratch	20 PB	Large and fast IO	Gadi only	980 GB/s	490 GB/s
/g/data1-5	80 PB	Storage of large data files	Global	450 GB/s	150 GB/s
Massdata	70 PB	Archiving data files	External	8 TB/hour	8 TB/hour

NCI - Cloud services

Nirin cloud computing platform

- 1856 high availability cores with 22TB memory
- 16640 high capacity cores with 32TB memory
- 40 GPUs in high capacity zone
- Openstack virtual machines



OzStar and beyond – Swinburne University

OzStar (since 2018)

- 4140 CPU cores and 230 GPUS in total spread over 107 “standard” and 8 high-memory compute nodes
- 13PB Lustre ZFS Target File System

Ngarrgu Tindebeek (“Knowledge of the Void”) – commissioned 2023!

- 11,648 CPU cores and 88 GPUs in total spread over 160 “standard”,10 high-memory and 22 GPU compute nodes
- Sees the same storage as OzStar

Institutional Computing

- Very different setups at each Australian University
  - Some have on-premise clusters (e.g. UNSW Katana)
  - Some have cloud only computing
  - Some have a hybrid model
- Usually only for staff of the university

Data Central Science Platform

- Astronomy dedicated facility
- 16 compute nodes: 8 database, 7 Kubernetes, 1 R-Studio/R-Shiny
- 2 high-memory (1TB) machines for web/API and Cloud services
- 8 GPUs across two servers
- 1PB storage
- Applied for 12 new compute nodes, 1PB extra storage, outcome expected soon

Data Central Science Platform

- Currently the **only** Astronomy Science Platform in Australia
- JupyterHub recently deployed – complex access control requirements
- Carta for visualisation
- R-Shiny and R-Studio service
- Remote desktops via Apache Guacamole – support Linux and Windows VMs
- Interested in migrating some services to NCI – already hosting archive services

A complex access model

- Pawsey, NCI & OzStar
  - Need to apply for VM/Kubernetes, merit allocation for HPC
  - Pawsey has dedicated resources for SKA pathfinders
- Institutional compute
  - Cannot easily deploy jobs without institutional access
- Data Central
  - Astronomy dedicated
  - Priority given to Australian-led projects, but open to discussion

Final thoughts

- Access to Australian HPC facilities is complex
  - Usually need to apply for compute power and HPC
  - Sometimes need to be staff at institute
- Conversations started to connect HPC facilities
- Need a coordinated effort to bridge the gaps and challenge the access models
- **Technical problems are solvable, but need to work on sociological issues**