HPC Cluster Hardware

CPU Cluster (2 login nodes, 17 compute nodes)

All servers have:

- Two (2) Intel Xeon Gold 6148/6248 CPUs - total of 40 cores @ 2.3GHz.
- Twelve (12) 32GiB ECC DIMMs
- Two (2) 480GiB M2-SSDs
- Two (2) 3.84TiB SSDs
- Two (2) SFP28 ports
- RJ45 port for shared BMC/system connectivity

Login nodes have an extra two (2) SFP28 ports with SFP+ optics installed for public network connectivity.

The end-of-life date for these servers is Aug-2023 (15) and Aug-2024 (4).

GPU/Graphical workloads

The current GPU nodes reach end-of-life in April-2021. They will either be eWasted or repurposed.

The next generation of GPU hardware will be delivered as a virtualization platform, delivering high-end graphics capabilities as well as virtualised accelerator capabilities. The latter will deliver capabilities that allow researchers involved in development and testing of workflows. The initial purchase will only involve three GPU cards, with upgrades only likely if high-end graphical workload requirements increase. JCU is likely to continue to purchase access to GPU resources from UQ (via QCIF) for computational GPU workloads.

Network Connectivity

Red text is used to identify connectivity that isn't redundant. All connectivity is ethernet/IP based.

	NFS	Mgt.	Public
Login nodes	50Gb/s	1Gb/s	10Gb/s
Compute nodes	50Gb/s	1Gb/s	

The internal build for login/compute nodes is *high density* which is why there is no redundancy for management network connectivity.

The JCU HPC cluster has been built with a 30:1 blocking regime - i.e., the interconnect between switches is ~1/30th of that which is required for tightly couple MPI workloads. This is one of the reasons why our job management system has been configured with an aim to reject all (not run any) multi-node MPI jobs.

Storage Infrastructure

Researchers do not have direct access to the systems/devices mentioned below. Redundancy exists for all storage connectivity - physical connectivity is double the max. bandwidth figures mentioned below.

	Provisioned capacity	Max. bandwidth	Filesystem(s)
DELL SC4020 Storage	600TiB	32Gb/s	N/A (block storage)
DDN SFA7990E Storage	516TiB	200Gb/s	`/gpfs01`

A DELL R640 server is connected to DELL SC4020 storage to deliver the `/home`, `/scratch`, & `/sw` filesystems to CPU/GPU resources.

Another DELL R640 server has been configured to manage the allocation relationships between our medici cache (`/gpfs01`) and the home node at QCIF.

NOTE: Researchers with a ARDC/QCIF allocation should be aware that your allocation is housed at QCIF. The medici platform at JCU has been purchased and configured to allow filesystem like access to your data at QCIF from our HPC cluster hardware. If all data held at QCIF were to be uncompressed, we'd have over 2PiB of research data there - our medici service can only cache about 20% of that (free space requirements must be met) at any given time.

Redundancy Notes

There is no server level redundancy built into HPC server infrastructure (NAS server, medici server, login nodes, compute nodes, GPU nodes). When a node fails, all jobs running on the failed node will be killed.

There is no network switch level redundancy available for HPC (private or public) at JCU. Switch level redundancy exists for connectivity to the DELL SC4020 storage platform.

Redundancy/Resiliency is built into individual storage platforms, but should a platform fail most/all associated services could be unavailable for weeks /months.