NSW Supercomputers Water Applications 2023

NSW government organisations share their knowledge and experience using supercomputers for water related data-intensive and complex modelling applications.

Introduction

The NSW Modelling and Monitoring Hub (MaMH) is a collaboration and knowledge sharing network across NSW government agencies, state-owned corporations and local government. It develops, captures and disseminates specialist water modelling and monitoring information supporting effective water, catchment and coastal management in NSW.

Information, resource and expertise sharing through the MaMH will benefit the government beyond the resources available within individual agencies. Knowledge is being developed and shared through the delivery of a portfolio of special projects resourced through the knowledge hub. Refer to www.mamh.nsw.gov.au for more information.

This MaMH special project paper summarises identified NSW government supercomputer use in 2023 with a focus on water modelling.

Background

In today's world, scientific and engineering works can require exceedingly high-speed computations. This is where *supercomputers* can be used to help scientists and engineers. A supercomputer is a class of extremely powerful computer with a higher level of performance compared to a general-purpose computer. Supercomputers are commonly used for data-intensive and computation-heavy scientific and engineering purposes, such as:

- climate and weather forecasting
- astronomical observation
- quantum mechanics
- molecular modelling
- nuclear fusion research
- new chemical compounds (especially for pharmaceutical purposes)
- cryptology.

Also, supercomputers have found their applications in a variety of areas in water modelling including:

- optimization
- hydraulic and flood modelling
- wave modelling
- groundwater flow modelling
- stochastic model runs
- physical ocean/coastal process forecasting and hind-casting

Approach

NSW government is aiming to better utilise supercomputers in water modelling. This paper supports technical collaboration in water modelling and monitoring across NSW government organisations, through the promotion of access to supercomputer resources. Learnings from previous supercomputer projects are collated. It is expected that shared MaMH organisation team project applications in the future will provide greater access to limited supercomputer resources.

There are multiple advanced computer technology options available described in the following sections including their advantages and disadvantages. To be able to establish which option (or a combination) will be the best value for money in the long run as well as open up opportunities for sharing resources and more collaborations, the following considerations are required:

- Understanding the most common use cases.
- Computer operation and storage requirements.
- Where is data located?
- Is it easy to transfer or does the computer needs to be closer to where data is stored?
- What application environment is required?
- What are the reliability and operational requirements?
- Who needs to access the data?
- Backup and recovery requirements
- A high-level scoping of what will it cost to run those typical use cases at each facility

Scope of the project

The present project seeks to support NSW Government water organisations to better utilise and share learnings associated with the use of advanced computer technologies. The project scope is phased to suit available MaMH special project funding.

Phase 1 of the supercomputer pilot project included:

- Consult with key NSW government organisations including DCCEEW (MHL, Water Knowledge, EES), Office of the CSE, Sydney Water, WaterNSW, HunterWater on the use and experience with supercomputer projects;
- Review available supercomputer capabilities, resources, and contacts; and
- Document useful contacts and lessons learnt for the application of water industry supercomputer projects in a short project setup guidance flyer.

Next phase of the MaMH project could include

- Establish joint supercomputer access and develop workflows for running modelling on the computing platform; and
- Pilot a joint organisation project.

Increasing demands for supercomputer services and advancement of the technology led to fast changes into the supercomputer's capabilities. This report documents the latest available information at the time of publication (February 2024).

Findings

Advantages and disadvantages of using supercomputers

Table 1 outlines the advantages and disadvantages of using supercomputers.

Table 1: Advantages and disadvantages of using supercomputers

Advantages	Disadvantages
Better computational power: ability to run complex models such as Monti Carlo approaches in a considerably short time.	Highly competitive prioritisation categories to access the facility.
Fewer memory and storage issues: extra memory and super-sized storage capacity.	Expensive: renting a supercomputer can cost more than $1000 for an hour due to the high costs to build, maintain and run the supercomputer.
Time-efficient: ability to run very long-time span models in a short time. As an example, a groundwater flow model run-time was reduced from two months on network of five computers down to 24 hours on Pawsey's supercomputer.	Such systems are not typically on the corporate ICT network.

Advantages	Disadvantages
Flexibility enables unprecedented project speed, agility and scalability.	Using Linux operational system. Direct Windows support is not available yet.
Cost-effective: using supercomputers can save a lot of time in managing tasks. This not only results in efficiency, it eventually lowers cost in the long run.	Each machine has a different procedure to upload, transfer data and troubleshoot which requires training in that regard.

Supercomputer categories

Supercomputers can be categorised as:

- Government-funded supercomputer infrastructures (capital refresh funded by Commonwealth) including:
  - National Computational Infrastructure (NCI): The facility is located in Canberra, a partnership between BOM, GA, ANU, and CSIRO. NCI's newest supercomputer is **Gadi** which replaced its predecessor Raijin in January 2020.
  - The Pawsey Supercomputing Centre is a government-supported high-performance computing national facility located in Perth, Western Australia. Pawsey supports researchers in Western Australia and across Australia through the Pawsey Centre. Pawsey receives operational and capital funding from Western Australia and Australian governments. Pawsey supercomputing centre includes Setonix, Nimbus, and Garrawarla. Garrawarla provides supercomputing services specifically in astronomy research. Note that the Pawsey operated Magnus (activity period: 2013-2022), Galaxy (activity period: 2013-2022), Zues (activity period: 2014-2022) and Topaz which were all decommissioned in 2022.
- Commercially provided Cloud-based services
  - Amazon Web Services (AWS)
  - Microsoft Azure Cloud Computing Services
- NSW State-funded HPC infrastructures
  - Science Data Compute (SDC)
- High-end GPU computers and clusters: High-Performance Computing (HPC) clusters consist of multiple interconnected computers that perform calculations and simulations in parallel, allowing for faster and more efficient processing of large amounts of data. A cutting-edge opportunity presents itself with the availability of high-end GPU clusters for purchase. These clusters offer substantial benefits, particularly for tasks requiring significant computational power and reliable support. With the ability to self-manage priorities, users can efficiently

allocate resources to meet their specific needs. However, the complexity of these systems necessitates ongoing support to ensure optimal performance and maintenance. Additionally, considering a shared organisation option could be advantageous, especially if there is a collective demand for these powerful computing resources.

Table 2 summarises readily available key technical specifications of the supercomputers accessible by NSW government.

Table 2: Technical specifications of supercomputers available for NSW government water organisations

Category	Vendor	Supercomputer	Architecture	Operating system	Technical specifications	User guide	Computing price
Government-funded supercomputer infrastructures	NCI	Gadi		Rocky Linux 8	• 10 login nodes, 6 data mover nodes, 4000+ compute nodes including 160 GPU compute nodes that also provide 640 NVIDIA V100 GPUs • There is 1.67 PiB storage available locally on compute nodes and also a 22 PiB Lustre parallel filesystem • 3,074 nodes each containing two 24-core Intel Xeon Scalable 'Cascade Lake' processors and 192 Gigabytes of memory; including 50 nodes each offering 1.5 Terabytes of Intel Optane DC Persistent memory • 720 nodes each with two 52-core Intel Xeon Scalable 'Sapphire Rapids' processors and 512 Gigabytes of memory • 804 nodes each with two 14-core Intel 'Broadwell' processors • 192 nodes each with two 16-core Intel 'Skylake' processors • 160 nodes each containing four Nvidia V100 GPUs and two 24-core Intel Xeon Scalable 'Cascade Lake' processors. • 10 nodes each with two 14-core Intel 'Broadwell' processors and 512 Gigabytes of memory • 2 nodes of the NVIDIA DGX A100 system, with 8 A100 GPUs per node.	Gadi User Guide	Cost
	Pawsey	Setonix	HPE Cray EX	CrayOS (based on SUSE Linux Enterprise Server)	• Total of 10 login nodes, 2x AMD EPYC 7713 "Milan" CPU, 2x 64 cores per node and 256 GB RAM per node • Total of 1592 CPU computing nodes, 2x AMD EPYC 7763 "Milan" CPU, 2x 64 cores per node and 256 GB RAM per node • 154 GPU computing nodes, 1x AMD optimised 3rd Gen EPYC "Trento" CPU, 1x 64 cores per node, 8 GCDs (from 4x "AMD MI250X" cards, each card with 2 GCDs), 128 GB HBM2e and 256 GB RAM per node • 38 GPU High mem nodes, 1x AMD optimised 3rd Gen EPYC "Trento" CPU, 1x 64 cores per node, 8 GCDs (from 4x "AMD MI250X" cards, each card with 2 GCDs), 128 GB HBM2e and 512 GB RAM per node • Total of 16 data movement nodes, 2x AMD 7502P CPU, 1x 32 cores per node with 128 GB RAM per node	Setonix User Guide	Service Unit
		Nimbus	Cloud operated	Limited online information available, contact help@pawsey.org.au			
Commercially provided cloud-based services	Amazon	Amazon Web Services (AWS)	Cloud operated	Linux/ Windows	Given the wide range of the instances available, it is highly recommended to check AWS website for details.	Here	Computing price
	Microsoft	Microsoft Azure	Cloud operated	Linux/ Windows		Here	

Category	Vendor	Supercomputer	Architecture	Operating system	Technical specifications	User guide	Computing price
NSW State-funded HPC infrastructures	NSW Government	Science Data Compute (SDC)		Linux	• Total of 13 compute nodes including CAS project with 4 servers and IRS project with 9 servers which consists of 2 GPU nodes as documented below: – SDCcomp01 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 256GB RAM – SDCcomp02 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 256GB RAM – SDCcomp03 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 256GB RAM – SDCcomp04 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 256GB RAM – SDCcomp05 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 256GB RAM – SDCcomp06 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 512GB RAM – SDCcomp07 with CPU of Xeon E5-2690 v2 @ 3.00GHz *2, AVX Ivy Bridge instruction set and 512GB RAM – SDCcomp08 with CPU of Xeon Gold 6248 @ 2.50GHz *2, AVX-512 Cascade Lake instruction set and 768GB RAM with Nvidia Tesla V100-PCI E-32GB Extras – SDCcomp09 with CPU of Xeon E5-2697 v4 @ 2.30GHz *2, AVX2 Broadwell instruction set and 256GB RAM – SDCcomp10 with CPU of Xeon E5-2697 v4 @ 2.30GHz *2, AVX2 Broadwell instruction set and 256GB RAM – SDCcomp11 with CPU of Xeon E5-2697 v4 @ 2.30GHz *2, AVX2 Broadwell instruction set and 256GB RAM – SDCcomp12 with CPU of Xeon E5-2697 v4 @ 2.30GHz *2, AVX2 Broadwell instruction set and 256GB RAM – SDCcomp13 with CPU of Xeon Gold 6248 @ 2.50GHz *2, AVX-512 Cascade Lake instruction set and 768GB RAM with Nvidia Tesla V100-PCI E-32GB Extras • 2 development nodes including irsdev01 and irsdev02 for users direct access	To access SDC user guide contact the MaMH Working Group	No charge
High-end GPU computers and clusters	-		HPC cluster	Linux/ Windows	Approx. entry-level cost in the $10,000 range up to $100,000 range for a high-end cluster		

Supercomputers in use by NSW government

NSW government organisations have used supercomputers to undertake data-intensive and heavy computations for years, including:

- DCCEEW Water Knowledge (Water Analytics - Modelling team) has used NCI's supercomputer Gadi and Pawsey including Nimbus, Zeus, Magnus.
- DCCEEW Environment and Heritage Group's Science, Economics and Insights (SEI) Division (Coastal and Marine Team) uses the NSW funded SDC to run state-wide wave models. This includes hind-casting (i.e. looking at past wave conditions), as well as forecasting NSW wave conditions to better understand how waves interact and impact coastal environments in NSW.
- EHG Science Economics and Insights division (Remote Sensing and Landscape Science branch and Climate and Atmospheric Science (CAS) branch) regularly use the SDC HPC for modelling applications.
- WaterNSW utilised AWS, Azure and High-end GPU computers such as gaming PCs.

MaMH organisations developed two case studies and a potential demonstration project to showcase examples of utilising supercomputers in water modelling:

- Utilising SDC in NSW Nearshore Waves modelling to provide context to coastal councils for smaller scaled detailed coastal hazards mapping and provide improved science on coastal processes to underpin the development of Coastal Management Programs (CMPs). By leveraging the computational power of supercomputers, the project has achieved accurate calibration, extended hind casting / modelling for the whole of NSW and enhanced decision support capabilities.
- Utilising SDC in groundwater flow modelling to minimise damage to key aquifers, support water resource management, inform water management policies, and optimise water usage. By leveraging the computational power of supercomputers, the project can achieve accurate calibration and enhance decision support capabilities, efficiently run complex groundwater models and deliver timely results.
- Hydrodynamic model of the Wallis Lake and estuary system using SDC to refine the understanding of sediment transport processes around Wallis Bridge, assess navigability, and investigate the potential impact on oyster leases. The proposed approach involves model refinement, calibration, and validation, followed by the analysis of various management options. By leveraging the computational power of supercomputers, the project aims to achieve accurate calibration, and provide efficient and scalable simulations. The project

requires dedicated computational resources to efficiently run a complex hydrodynamic model and deliver timely results.

Prioritising supercomputer applications

Increasing demands for supercomputer services have led to competitive allocation schemes for universities and agencies across Australia to decide access to supercomputer resources. Methods for accessing supercomputer systems and services are described as follows.

For NCI:

- Access Schemes including:
  - National Computational Merit Allocation Scheme (NCMAS): The main call for applications is made annually around August for allocations to start the following January for 12 months.
  - Adapter Scheme is designed to identify meritorious research projects that need flexible compute access over a short period. It provides quarterly allocations up to 250KSU to Australian researchers, catering for those users who need smaller compute and data resources in short bursts.
  - AI Flagship Scheme provides access to cutting-edge AI infrastructure at NCI, including GPUs, CPUs and specialised AI hardware accelerators.
  - NCI Start-up Scheme is available to researchers who would like to evaluate the suitability of NCI for their research and/or to assist in framing applications for more substantial grants. Start-up grants are allocated on NCI's supercomputer at a maximum of 5KSU per year.
- Collaborating Organisations: around 75% of NCI's HPC resources are reserved for researchers at their collaborating organisations. These organisations generally distribute allocated HPC resources through internal processes. Researcher or employee at one of NCI's collaborating organisations, can gain access to NCI services via contacting their Scheme Manager.
- NCI Flagship Allocation Scheme: This scheme provides access to HPC, data-intensive and storage services at NCI for projects identified by the NCI Board as being of high-impact or national strategic importance.
- Virtual Research Environment (VREs): VREs bring together a community of researchers across multiple organisations to enhance collaboration on national and international science priorities. Visit Virtual Research Environment for further information.
- Commerce and industry access on a fee-for-service basis. Contact help@nci.org.au for details.

For Pawsey:

- Project allocations are awarded via a competitive merit process (similar to NCI). There are two access mechanisms, driven by the *Access Principles* of the National Innovation Roadmap (Here).

  - Merit: the bulk of supercomputer access at Pawsey is through competitive merit.
  - National Interest: at Pawsey, some infrastructure is reserved for the precursor telescopes of the Square Kilometre Array, ASKAP and MWA.
- The following access schemes are available:
  - The National Computational Merit Allocation Scheme (NCMAS) (annual call in September/October each year)
  - The Pawsey Partner Merit Allocation Scheme (annual call in September/October each year). The Partner institutions are CSIRO, Curtin University, Edith Cowan University, Murdoch University and The University of Western Australia.

For both NCI and Pawsey:

- Monash University allocation scheme - The Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) funding for the other users - competitive allocation runs for universities and research agencies across Australia.

Funding

The current supercomputer applications are being funded through:

- Commonwealth government funds major capital refreshes for NCI and Pawsey. Other costs are funded by partner shares for NCI which decides resource allocation to these partners.
- SDC is funded by DCCEEW from their program's funding.

Summary

This paper summarises available information regarding supercomputers accessible by NSW government organisations and through which pathways. Based on the case studies undertaken in this study, it is evident that supercomputers play a crucial role in advancing water modelling capabilities for NSW government organisations. The successful utilisation of supercomputers in projects such as the NSW Nearshore Waves and groundwater flow modelling demonstrates their effectiveness in addressing complex challenges. These case studies highlight the significant benefits of leveraging advanced computing technologies, including improved accuracy, efficiency, and scalability in modelling efforts. Furthermore, they underscore the importance of collaborative initiatives like the NSW MaMH in facilitating knowledge sharing and maximising the impact of limited supercomputer resources.

Moving forward, it is imperative for NSW government organisations to continue exploring opportunities to harness the full potential of limited supercomputers access in water modelling

applications, through sharing learnings from previous supercomputer projects and combined project applications. It is recommended for the MaMH to:

- Undertake further investigations on HPC clusters;
- Investigation of the costs of developing HPC clusters for NSW government;
- Further investigating supercomputers with Windows architecture;
- Further investigate and promote the capabilities of utilising SDC;
- Identify opportunities and trail supercomputers for hydraulic modelling. Hydrodynamic model of the Wallis Lake and estuary system using SDC is suitable and ready for trail application. Determine the value for money of the application.
- Further investigate the actual costs of utilising supercomputers.