National Supercomputing Centre

A Guide to HPC Resource Estimation

NSCC.SG

HPC Resource Planning – Key Takeaways

1. **Use past performance data for estimation whenever possible**
   - However, be aware of potential issues: outdated applications (not optimised) and older hardware may lead to inaccurate predictions.
   - Ensure you are using the latest software versions and optimisations. If unsure, read the documentation or consult us.
2. **Understand the HPC hardware and software stack.**
   - System architecture impacts performance—refer to documentation, attend training, or seek guidance if needed.
3. **Estimate your required resources accurately.**
   - Compute CPU core-hours, GPU card-hours, and storage needs based on your application's requirements.
   - If running multiple applications, estimate each separately before summing them up for your final request.
4. **Include a buffer in your resource requests.**
   - Allocate at least 10% additional resources to accommodate debugging, restarts, and unforeseen issues.
5. **All users must attend the 2A/2A+ introductory workshops.**
   - Users who skip these workshops often experience resource wastage, sometimes up to 100%, due to unused CPU/GPU allocations.
   - Proper training ensures efficient use of resources and minimises waste.

HPC Resource Planning – The Don’ts

1. **Requesting Resources Without Testing**
   - Unjustified resource requests can be flagged as unrealistic
   - Example: Requesting GPU resources for genome sequencing without checking if the bioinformatics tool actually supports GPUs.
2. **Skipping Workflow Development & Benchmarking**
   - A structured workflow improves efficiency and ensures optimal resource use.
   - Benchmarking helps you understand your workload for better planning and potential improvements.
   - Even a simple benchmark and workflow help
3. **Making Imbalanced Resource Requests**
   - Example: Requesting 10 GPU card-hours but 1000 TB of storage.
   - HPC systems are designed for computation, not large-scale cloud storage.
   - Plan resource requests proportionally to your actual compute and storage needs.
4. **No Human Resource Planning and Handover**
   - Ensure staff continuity for workload execution and transition plan to prevent disruptions.
   - We have seen cases where key personnel leave without proper handovers, causing long delays.

Useful Links for Resource Estimation

Nvidia Benchmark (AI)– GPU

- https://developer.nvidia.com/deep-learning-performance-training-inference/training
- https://developer.nvidia.com/deep-learning-performance-training-inference/ai-inference
- https://developer.nvidia.com/deep-learning-performance-training-inference/conversational-ai
- https://docs.nvidia.com/nemo-framework/user-guide/24.09/performance/performance_summary.html

AI Training or Fine-tuning Resource Guideline – GPU

- DeepSeek-V3 Technical Report
- https://docs.api.nvidia.com/nim/reference/meta-llama-3_1-405b
- Understanding the Performance and Estimating the Cost of LLM Fine-Tuning
- https://en.wikipedia.org/wiki/Neural_scaling_law

Nvidia Benchmark (HPC Applications) – GPU

- https://developer.nvidia.com/hpc-application-performance

GROMACS, AMBER, NAMD, LAMMPS (HPC Application) – CPU and GPU

- https://www.hecbiosim.ac.uk/access-hpc/our-benchmark-results/archer2-benchmarks
- https://www.hecbiosim.ac.uk/access-hpc/our-benchmark-results/bede-benchmarks

National Supercomputing Centre

Thank You

NSCC.SG