

PROJECT GAUSS

Data-Driven Scheduler Optimisation for Supercomputing Systems



How To Apply

Send your CV, academic transcript and a brief introduction to recruit@nscg.sg with the subject line: **Internship Application - [Your Name] [Project Title]**

Please include your intended internship period within the email. Our team will be in touch with shortlisted candidates.

The Challenge

The National Supercomputing Centre (NSCC) Singapore manages Singapore's national high performance computing (HPC) resources, a critical asset that drives the advancement of cutting-edge research, technology, and innovation.

While hardware dictates theoretical peak performance, the actual realised throughput of a supercomputer is heavily influenced by how effectively the job scheduler orchestrates diverse workloads. As our user base grows, we are advancing our analysis of historical job logs to uncover subtle opportunities for optimisation.

We aim to conduct a systematic deep-dive into this data to provide the empirical evidence needed to model next-generation scheduling policies in a simulated environment.

What You Will Do

As part of this optimisation initiative, you will contribute to one of two key pillars of supercomputing efficiency:

- **System-Side Logic & Throughput:** Analyse the mechanics of the PBS scheduler to identify technical opportunities for higher throughput. You will investigate variables like backfill efficiency and queue contention to pinpoint how we can further minimise idle compute time through refined scheduling parameters.
- **Workload Characterisation & Service Design:** Study population-level trends in how the research community utilises the system. By analysing patterns such as wall-time request accuracy and submission seasonality, you will help us design more responsive, user-centric services that align with actual workload requirements.

What You Will Deliver

A functional POC for your specific module—such as a simulation component or a targeted behavioural analysis report. Your findings will help refine future scheduling strategies, ensuring NSCC continues to provide a world-class, high-throughput environment for Singapore's researchers.

Required Qualifications

- Currently pursuing a degree in Data Science, Computer Science, Statistics, or a related analytical field
- Strong programming skills in **Python** (specifically Pandas, NumPy) or **R** for data analysis
- Solid foundation in **statistics** and processing large datasets to identify trends and anomalies
- Strong analytical problem-solving skills to translate raw data into actionable insights

Bonus / Preferred Skills

- Familiarity with operating system concepts, specifically process scheduling and resource allocation
- Experience with discrete event simulation or modelling complex systems
- Knowledge of HPC workload managers (e.g. PBS Pro, Slurm, LSF)