

CS Capstone: Policy and process for getting access to High Performance Computing Resources



Executive Summary:

Some projects doing machine learning, big-data modeling, or other particularly compute-intensive processes may have a legitimate need for access to high-performance computing (HPC) resources. NAU has a powerful cluster called Monsoon, but time on this machine is scarce and costly, so access can be granted only when truly needed.

HPC@NAU:

As the “big data” revolution continues and computing needs of much data science research (e.g., AI, modeling, etc.) have exploded, NAU has invested in a powerful HPC facility centered around a cluster called Monsoon. The size and capability of Monsoon are being updated/expanded all the time, but at the time of this writing, Monsoon has [2860](#) Intel Xeon cores, 24TB of RAM, and 20 NVIDIA GPUs. It short: pretty big-bore compute power! The HPC group is embedded within the larger ITS infrastructure at NAU, with several staff members in the group, led by Chris Coffey. For up-to-date info on Monsoon hardware, performance, staff, and access policies, see the HPC website:

<https://in.nau.edu/hpc>

How to get access: policy and procedure

HPC access is a valuable and heavily-subscribed resource. Even NAU faculty have to request access, have a limited amount of “HPC time” allotted to their projects, and this time is accounted and paid for by their grants and/or departments.

Due to the increasing number of AI, machine learning, and modeling projects being proposed by Capstone clients, we have created a special charging account for CS Capstone access to Monsoon, as well as a procedure for getting Capstone teams with demonstrable HPC need access to the Monsoon resource. In general, the steps are:

1. Think carefully about why your project will need HPC access, and when that access will become relevant to have.
2. Meet with and convince the Capstone faculty organizer (Dr. Doerry, currently) that your project needs Monsoon access. Describe your need and at exactly at which point in the project you anticipate needing it.
3. When given the go-ahead, follow the HPC group’s guidelines for requesting Monsoon access. As part of application processing, Dr. Doerry will indicate his approval for your account.
4. Schedule Monsoon training. All Monsoon users must have training to understand how the cluster works and how the SLURM scheduler works to get your job run on the cluster. These trainings happen periodically throughout all semesters; see the HPC website for upcoming sessions.
5. Receive your activated account and use Monsoon.

The first step in this process is the most important by far. Whatever HPC need you have in your project, *it’s probably relevant only near the last segment of your project*. To do your development and testing work – most likely on small “toy” data and datasets -- you should install whatever compute-intensive software you’ll be using locally, on your own machine first, so that you can have full control and convenience in working out the code. Only when you have your software refined and are ready to move on to larger problems and datasets should you request Monsoon access. Be ready to clearly document

growing processing power needs (e.g., this already now takes xx hours overnight to run on my laptop, and we expect that to double when we move to real client datasets. Need more power!) before you go and see Dr. D. In short, you'll need to convince Dr. D that you are ready to go and really need HPC access before authorization will be granted. This is not meant to be a big hurdle and it won't be; it's just there to avoid wasting valuable HPC time running simple testing and dev jobs that could run locally.

Other important considerations:

- **Sustainability of your project after Capstone.** An obvious concern here is that, if your Capstone software product includes a central HPC element to operate, how will the client keep it operating after delivery? Strictly speaking, this isn't your problem...you've delivered the software built to spec, on-time. Best practices, however, would certainly require you to discuss this issue with your client early and often. Some less tech-savvy NAU clients just assume that, because they're at NAU, they have easy and unlimited access to Monsoon. This is not the case. Make sure your client understands that they will need to apply for access and direct them to the HPC group for assistance/advice.
- **Understand limited access channels to Monsoon.** As you can imagine, high-powered research computing is not something that you want to have poor security on. For this reason, access to Monsoon is quite tightly-controlled, i.e., you can generally communicate with Monsoon only over SSH in a terminal. Plus, once connected, you don't just run stuff; you need to schedule jobs using the SLURM HPC scheduler. What all of this adds up to is that it can be a challenge (or impossible in some cases) to integrate Monsoon into your product as an invisible HPC backend that you just send work to as needed behind the scenes. If this sort of integrated "transparent, anytime, with instant response" access is what your client needs or wants, consider a cloud-based (e.g. AWS) infrastructure that includes an on-demand HPC element. Of course, this would be much more costly than the (free) Monsoon access...but that's just part of the real-world computing equation.