Supercomputing In Plain English

What's a supercomputer?

A supercomputer is a collection, or cluster, of a large number of regular computers (referred to as compute nodes) connected over a network. Each of the computers is like a local workstation though typically much more capable. For example, a standard laptop might have 4 CPUs and 8gb of RAM. Compare this with a standard compute node on Puma which has a whopping 94 CPUs and 470gb of RAM!

Another thing that differentiates supercomputers from your personal workstation is a supercomputer is a shared resource. This means there may be hundreds or even thousands of simultaneous users. Without some sort of coordination, you can imagine it would be a logistical nightmare to figure out who can run their code, what resources they can use, and where they should run it. That's why supercomputers use login nodes and job schedulers (we use one called SLURM).

When you first log into a supercomputer, you're connect to something called a login node. This is a single computer that's connected to the cluster of compute nodes and is where you write and submit your jobs using a job scheduler. A login node itself is not used for performing any analyses.

A job scheduler is software used to coordinate user jobs. You can use it by writing a special script that requests compute resources (e.g., CPUs, RAM, GPUs) and includes instructions for running your code. You submit this script to the job scheduler using a special command called sbatch and it does the work of finding the required space on the supercomputer for you. It then runs your code and returns the results to your account in a text file.

One way you might think of a supercomputer setup is as a post office and factory. Imagine you have something you need built in a factory and have a list of instructions and materials for how to do it. To achieve this, you put your instructions (code) in an addressed envelope (SLURM script), take it to a post office (login node), have postal worker (job scheduler) deliver the instructions to the factory (compute node), and then you can go home (log off). After waiting for a period of time, your completed project is delivered to you.

One of the computers, or compute nodes, in Ocelote. They have the same components as a PC or workstation: there are two processor modules, memory DIMMs, an internal disk and networking ports. The power supplies are in the chassis that the compute nodes plug into. Some of the compute nodes have GPUs like the ones in your Xbox or gaming laptop that you use for Minecraft but are much more capable.
Why should I use one?

There are lots of reasons to use a supercomputer! For example, say you have analyses that require a tremendous amount of memory or storage space. It's not feasible (or very expensive) to use 3TB of memory or 10TB of disk space on a local workstation, but on our systems it's very possible (and free). This is how you scale up from the workstation under your desk.

Another possibility is you may have thousands of simulations to do. This may take an unreasonable amount of time and be a serious bottleneck for your research if you’re running them in serial locally. However, on a supercomputer you can run hundreds of jobs at the same time using thousands of CPUs. This means you may wind up getting results in hours instead of months. This how you scale out your work.

You may also have experience with being frustrated with a job’s runtime. What happens if it takes a week or longer to complete one of your analyses? On a local workstation, keeping your computer awake for the duration of the run may be difficult, inconvenient, or impossible. On a supercomputer, the process of running jobs can be fully automated. Once you have a special script written with all the necessary instructions, you can submit it to the scheduler and it does the rest. This means you can log out and close your computer without any worry about interrupting your work. Your results are returned to you as a text file in your account in real time so you can always log in and check your progress. You can even request email notifications to keep track of your job’s status, though you’ll want to be careful not to mail bomb yourself if you’re running thousands of jobs.

How do I access a supercomputer?

Though you’re physically distant from the UArizona supercomputers, you can access them directly from the comfort of your own local workstation. This can either be done from a local terminal (for Windows users, using something like PuTTY) or by using our web interface. To begin the process, you’ll first need to request an account.

Where do I store my files?

Every compute node in our supercomputers is connected to a large storage array. This storage array is where any files you upload to the system are saved and means that no matter where you are on the system, you’ll be able to access your data. There are three locations where your files may go, each with a different size limit. Take a look at our storage page for more detailed information.

What about software?

We provide hundreds of software packages as modules. These can be loaded using special module commands.

Are there any resources available to help me get started?

Absolutely. We have a quick start tutorial that can help walk you through the specifics of logging in and submitting a job for the first time. We have an Intro to HPC workshop video recording that goes over the basics of supercomputers and how to use them. We also have consultants available to provide support. They can be reached either via a ServiceNow ticket or by dropping by our virtual office hours held every Wednesday from 2:00-4:00pm. Click here for information on user support.