Resource Amount Selection

We like to see our usage statistics as high as they can be. Our ideal setup would be 100% utilization with little wait time. Realistically that won't happen. One of the largest factors that impacts utilization is when a user requests resources they don't use; for example: job requests 16 cores but is single threaded. The other 15 cores would be locked out from other users and hold up the queue. We're a bit more flexible with RAM as it's much harder to estimate usage before running your jobs.

Resource utilization summary:

- RAM
    - Under request: Job will die when it runs out of RAM
    - Over request: Grumpy admins will reach out and try to get you to lower your request.
    - Ideal: over request but try to keep wasted RAM < 10% of your request. Closer to 0% wasted the better.

- Cores
    - Under request: Your job will step on itself because of kernel scheduling. The job will take a massive performance hit as a result.
    - Over request: We reserve the right to kill your job on sight.
    - Ideal: Request exactly the number of cores your job will use.

- Time
    - Under request: Your job WILL die when it runs out of time.
    - Over request: Slurm may deprioritize your job to let smaller jobs go first. May not fit within scheduled maintenance windows. (Run the command `time-until-maintenance` to see scheduled maintenance windows.)
    - Ideal: Over request by try to keep your time as close as possible.

To see all available resources, use the command `cluster-free`. Unfortunately, this will show some computers that may not actually be available for your use. The ones that are available are listed in the command `sinfo` under the partition 'work'. The maximum size for a single node job is currently 60 cores, which will run on the computer overkill. It may take you a bit of time to get that machine though, especially if there is a long queue. The cluster will attempt to prioritize smaller, faster jobs.