ICHEC Access for University of Galway MScAI / MScAI Online Oct/Nov 2022

James McDermott james.mcdermott@nuigalway.ie

This document is intended to get our students up and running with ICHEC supercomputing resources. It applies to the Oct/Nov 2022 sessions, with usernames `courseXXX` supplied by ICHEC.

Key concepts

- There are *three* machines in play: your local machine (your laptop), the Kay login node, and the Kay compute node you are using. Files are synced between your home directories on the Kay login node and compute node, via NFS, so you don't have to copy files. But there is no syncing between your laptop and the Kay login node, so you have to copy files to/from manually (see below).
- On ICHEC, you can do some setup interactively on the login node, and then when all is ready you *submit* a *batch job* which goes on the SLURM queue and executes in the background. You do not run large jobs, eg long-running Python scripts, directly.

Accessing Kay and running a batch job

This section is all you need for a basic batch job (no special Python libraries, no GPU).

1. Create your `ssh-key` and upload to the website provided by ICHEC (already done).
2. Find your `username` in the table provided in Bb. It is `courseXXX` where `XXX` is some digits. Your password is provided in Bb also.
3. Check you can login to a Kay login node (`kay.ichec.ie`) using `ssh`, `PuTTY`, `MobaXTerm`, or similar.

   You'll find yourself at a Linux prompt where you can use commands like `pwd` (print current directory), `ls` (list files in current directory), `mkdir` (make directory), `cd` (change directory), `mv` (move or rename a file), `rm` (delete a file), `less` (read a text file), and `python`. If necessary you can set up your Python environment here (see **Virtual Environments** below).

   ICHEC have stated we cannot change the password. But don't worry, these accounts are temporary and we don't anticipate any personal data being used.
4. Keep the login open, and use a text editor on your laptop to edit your `taskfarm.sh` file. Put in your own email address. Everything else can stay the same (for now). Take a look at `submit.sh` and `randmax.py` also as they demonstrate some useful ideas for scientific computing.
5. Now in a new window, copy your files from your laptop to your home directory on the login node. If using `scp` it will be something like the following, where `username` is replaced by your username.

   ```
   scp randmax.py submit.sh taskfarm.sh username@kay.ichec.ie:/ichec/home/users/username/
   ```

6. Now back to your Kay login. Using `ls` you should see that your files have appeared. You can run `sbatch submit.sh`. It should tell you that your job has been queued. You can now safely log out of Kay, close your laptop, etc.

7. When you receive an email to say the job has completed, you can log back in to Kay to look at the results files `randmax_0.txt` and the `slurm*.out` files. If there was an error, see **Troubleshooting** below. You can also copy these files back from Kay to your laptop to look at them there, store them, do statistics and plots. This is good practice in general, but particularly because our `courseXXX` usernames are temporary and the home directories will be deleted after our exercises are finished.

Virtual Environments

If your Python code needs to use extra libraries, we can use Anaconda to install them into a *virtual environment* on Kay.

1. Login to the Kay login node. First we tell Kay we are going to work on Conda:

   ```
   module load conda/2
   ```

   You will see a message like `To load the default (base) ...`. It is not an error! It is fine, just continue.

2. Create a new Conda environment called `myenv` and install some packages:

   ```
   conda create --name myenv python=3.7
   source activate myenv
   conda install scikit-learn
   ```

3. We don't need to re-install this every time, it is once-off. But now that this environment exists, you can tell your future batch jobs to use it by adding `module load conda/2` and `source activate myenv` to `submit.sh` (they are currently commented-out).

GPU Nodes

GPU nodes are NOT AVAILABLE during the October-November 2022 sessions with usernames `courseXXX`. We will see GPU later on in the Deep Learning module, and if you need GPU for your Capstone project, contact me to set that up.

Troubleshooting

1. If you didn't supply a correct `ssh-key`, you don't currently have access. If you still want to do these exercises, send an email to `support@ichec.ie` with the subject 'Att. Goar: MScAI course'

2. We'll try to troubleshoot some of the common problems during our lab sessions.

3. Several commands allow you to query the batch job system, while on `kay`.

   ```
   $ mybalance
   $ sinfo
   $ squeue # eg `squeue | grep username` # substitute your username
   $ scancel # use `man scancel` to read options
   ```

In particular, if you have submitted a job and it doesn't seem to have started, you can use `squeue | grep username` to look for it. Usually, you just have to wait: probably only a few minutes if you have only requested one node for a few minutes, or a day or two if you have requested more. But you might see `AssocGrpBillingMinutes` in the final column: this means that the project has exceeded its resource limits and your job may not run. E.g. on `nuig02`, it will not run until the next month.

4. After your job is finished, on Kay you should find a `.out` file, e.g. `slurm-643296.out`. This is plain text and you can read it with `less slurm-643296.out`.

   If your `.out` file says: `CommandNotFoundError: Your shell has not been properly configured to use 'conda activate'.`, please check that `submit.sh` uses `source activate`, not `conda activate`. ICHEC recommend to use `source activate`: https://www.ichec.ie/academic/national-hpc-service/software/python-conda. However, if you prefer `conda activate`, it will work if you use these two commands first (in your `submit.sh` script):

```
CONDA_BASE=$(conda info --base)
source $CONDA_BASE/etc/profile.d/conda.sh
```

   (This is from https://stackoverflow.com/questions/49600611/python-anaconda-should-i-use-conda-activate-or-source-activate-in-linux.)

   If your `.out` file says something like "unexpected end of file", it might because there is a blank line or a comment line in `taskfarm.sh`. It won't prevent the other lines from running correctly, but it's good to remove these so that it's easier to spot/troubleshoot other possible errors.

   If your `.out` file says something like `ImportError`, it could be because the environment was not created correctly. Sometimes it happens that your python is python 2 instead of python 3. Try replacing `python` in the `taskfarm.sh` with `python3`.

   If your `.out` file says something like `XDG_RUNTIME_DIR`, this can sometimes happen if you try plotting (using `matplotlib`) in your python code. It's better to copy your result files back to your laptop, and do any plotting there.

ICHEC documentation for further reading

- Quick start guide: https://www.ichec.ie/academic/national-hpc/documentation
- Tutorials https://www.ichec.ie/academic/national-hpc/tutorials
- Kay user guide https://www.ichec.ie/academic/national-hpc/kay-user-guide
- Python, Anaconda, and environments on ICHEC: https://www.ichec.ie/academic/national-hpc-service/software/python-conda
- Conda environments: https://www.ichec.ie/academic/national-hpc/documentation/tutorials/conda-environments
- More on SLURM: https://www.ichec.ie/academic/national-hpc/kay-documentation/slurm-workload-manager
- SLURM commands: https://www.ichec.ie/academic/national-hpc/kay-documentation/slurm-commands
- SLURM for PBS users (still has some useful clues): https://www.ichec.ie/academic/national-hpc/kay-documentation/pbs-slurm
- Task farming: https://www.ichec.ie/academic/national-hpc/documentation/task-farming
- Hardware available: https://www.ichec.ie/about/infrastructure/kay
- Software available (only relevant for hardcore HPC such as simulation, weather forecasting, etc, not relevant to those using the Python scientific stack): https://www.ichec.ie/academic/national-hpc-service/software

- Applying for your own *Class C* project, eg for a Capstone which needs GPU (talk to your supervisor and/or James McDermott): https://www.ichec.ie/academic/national-hpc/national-service-projects