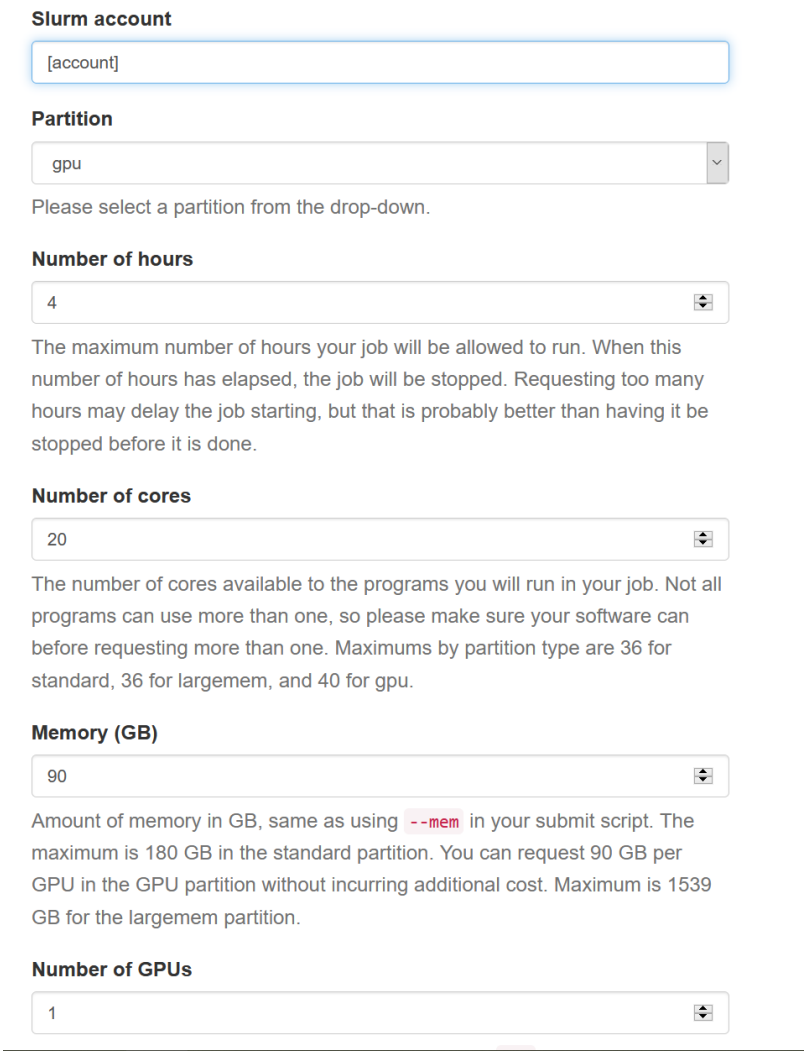
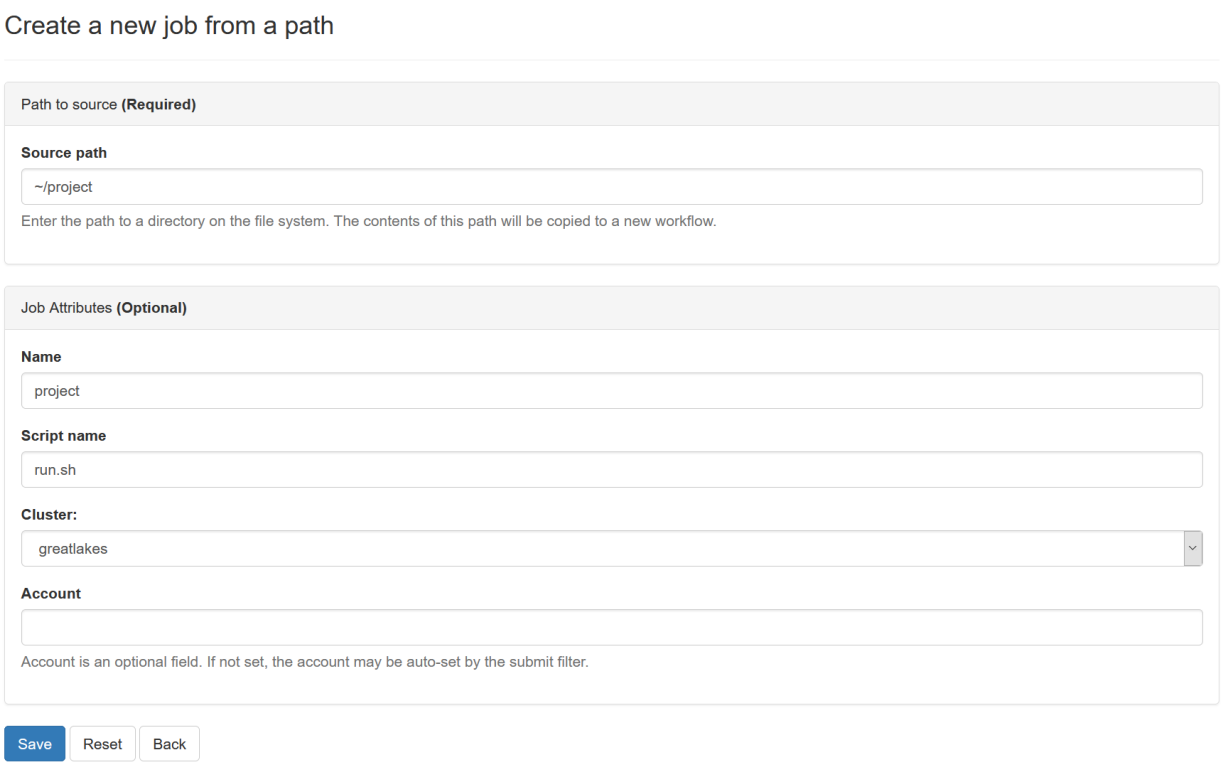
Great Lakes Tips

Using GPU Resources

There are two ways to use GPU resources on Great Lakes. One is just using the "Interactive Apps". For example, to launch a Jupyter Notebook server using a GPU, we could fill in the following options and launch the server. Note 1 GPU is enough to run the job, and as there are only 5 GPUs for a whole class, it's better to not use 2 GPUs at the same time.



The other one is to submit a Slurm job with specified submission options in the batch script. Besides using `sbatch` from command line, we could also compose a job. We may first compose "from the specified path". For example,



Later we could just compose "from the selected job". An example Slurm batch script `run.sh` is included in the `Project` folder, and for more usage, please refer to https://arc.umich.edu/greatlakes/slurm-user-guide/.

Some potential problems and the corresponding solutions

- Could not install simpletransformers

  Solution (According to ARC-TS):

  If you are using Jupyter Notebook (or Lab), any packages that you install on the command line needs to be done using the same version of python3.x-anaconda that you have selected for use with Jupyter Notebook in your Open OnDemand session.

  So, for example, if you are using the python3.8-anaconda module in Jupyter Notebook, when you install packages from the command line, you should do so as follows:

  ```
  1  module purge
  2  module load python3.8-anaconda
  3  pip3 install --user simpletransformers
  ```

  The `--user` tag will, by default, place packages in

  `$HOME/.local/lib/python?.?/site-packages`

  where ?.? indicates the versioning of the Python release. The library will then be available to you for this and future sessions.

- Waiting for the GPUs in queue

  Solution:

  There are 5 GPUs for the whole class, so start as soon as possible. In addition, the queueing order depends on number of cores and number of time, so changing these two options may have some effect.

- CUDA out of memory

  Solution:

  Reduce the batch size or the sequence length (e.g., by truncating songs to at most *k* words)

- Stuck during training or out of disk space

  Solution:

  Delete the unnecessary huge files.

  (According to ARC-TS) To see what files are taking up the most space you can run this command:

  ```
  du -S -h /home/$USER | sort -r -h | less
  ```

  Some unexpected huge files may be generated by deleting files in the Jupyter Notebook server, which will exist in `~/.local/share/Trash`, or the model checkpoints saved by Simple Transformers in the corresponding output folder. For the unwanted checkpoints, we could delete them manually, or set options in Simple Transformers models, for example, `save_model_every_epoch=False`.