Date: 01/25/2020

Name: David Goerig (djg53)

School: School of Computing

Programme of Study: Advanced Computer Science (Computational Intelligence)

Module name: C0890 Concurrency and Parallelism

E-mail: djg53@kent.ac.uk

Academic Adviser: DR P Kenny (P.G.Kenny@kent.ac.uk)

Executing Batch Jobs on a High Performance Computing Cluster

1. Challenges

The main challenge in this exercise was to understand how to submits multiple batch script to Slurm.
These batch script are launching the training of multiple datasets on the machine learning programme Weka.
Many steps have to be done:
- train each database with Weka
- compute all the results to obtain name and percent
- compute all the percent to make an average.

The challenge is to automatise all these steps, to don't run everything by hands (using bash) on the school cluster.

2. My conclusion

Bash is useful, easy to learn, but too rigorous on clerical errors (compared to python).
Sbatch is useful for the cluster, and using cluster is a really good things for such big calculation. Because you can easily choose the cpu / gpu, the maximum running time, etc, and having easily the outputs.
For the last steps, it was needed to compute all the steps again, but these time it was done very easily, just by launching three scripts, it took only few seconds.

3. What I've learned

First I learned bash and sbatch, i never used both of them. But also that in real time systems we don't have expectation time of how long the job is and what is estimated time to complete it. But in batch systems the processor knows how long the job is as it is queued.
The batch systems can manage large repeated work easily. Repeated jobs are done fast in batch, without any user interaction, just by launching scripts.