I/O Performance Analysis

Adrian Jackson
a.jackson@epcc.ed.ac.uk
@adrianjhpc

epcc

1. Aims

The aim of this exercise is to have a look at the performance of some I/O benchmarks we have run on a large HPC system (ARCHER), and compare performance to data from the filesystem logs. This is a pen and paper exercise, we won't be running any code, but looking at graphs that show data from both the benchmark application and the system.

The graphs are available online at, via git, i.e.:

git clone https://github.com/adrianjhpc/iotutorial/

or if you can't/don't want to check out the git repository you can view online at:

https://github.com/adrianjhpc/iotutorial/tree/master/exercises/analysis

We have three different sets, showing different performance. They are arranged in different folders, matching the three exercises outlined below. Each directory contains some pictures that we will use to look at performance and what was going on in the system at the time the performance was measured.

2. First set

In the first set (https://github.com/adrianjhpc/iotutorial/tree/master/exercises/analysis/firstset) we will look at the performance of MPI-I/O, NetCDF, and HDF5 over a 10 months on ARCHER. There are graphs showing performance of all three I/O libraries either using full lustre striping, or the default lustre striping (4 stripes). This is reflected in the name of the images. For instance, hdf5_4_stripe.png is the HDF5 benchmark using the default 4 lustre stripe count on ARCHER. hdf5_max_stripe uses maximum striping on lustre on the same system. We have also provided some *zoomed* graphs, that focus on a narrower part of the timings to enable some more detail to be viewed in the majority of the data points.

Each benchmark is run 10 times, with the mean, minimum, and maximum displayed on the graph. Each benchmark is run at a similar time as the others, once a day (although some days are missed). Whilst the benchmarks should run 10 times each day, if the I/O is slow not all benchmarks will finish. Particularly impacted by this are the NetCDF benchmarks as they are run last. This means that the difference between mean and minimum or maximum may not be as significant for these benchmarks as it is for the MPI-I/O benchmarks.

For these graphs consider the following questions:

- What impact does striping have for the three I/O libraries?
- What is the performance different between the three I/O libraries?
- What variation do you see over time?

3. Second set

In the second set (https://github.com/adrianjhpc/iotutorial/tree/master/exercises/analysis/secondset) we focus on benchmark results for a particular month on the system. We still provide graphs for the different I/O libraries and striping, however, we now also provide some graphs that show the filesystem throughput and metadata server load for the same period.

The system graphs show data from all nodes, from just the compute nodes, from the login nodes (esl), and the post-processing nodes (espp).

For these graphs consider the following questions:

- Is there any correlation between the variation in I/O performance and what is going on in the system as a whole?
- Is it I/O through-put or metadata operations that are causing issues?
- Are the impacts on the different I/O libraries different?

4. Third set

The third set (https://github.com/adrianjhpc/iotutorial/tree/master/exercises/analysis/thirdset) follows on from the second set. There are the same types of graphs, but from a different month. Have a look at them and see if there are differences from the second set? Are there different causes for the I/O issues?