Contribution ID: **26**   Type: **Full presentation: long (25 mins.)**

Using OpenStack to share hardware between Big Data, AI, HTC and HPC workloads

Wednesday, 4 November 2020 13:15 (15 minutes)

At StackHPC we work with many public and private institutions to build clouds that work well for their Scientific Computing needs. At Cambridge University, we have helped build their new Arcus cloud. It supports VMs, Containers and baremetal instances within a single cloud. This enables a diverse set of communities to share a single pool hardware resources including Kubernetes based environments (such as JupyterHub, Kubeflow and Pangeo), traditional batch job HPC clusters (typically Slurm with low latency networking) and allowing science communities to consume infrastructure directly and run their own custom science platform. This is all powered by the ongoing convergence of hardware needed by these various workloads.

In this talk we look back at the lessons Cambridge University have learnt over the years running a wide variety of workloads across OpenStack and Slurm. We then take a detailed look at how they are currently using to provision all of their new baremetal servers, rather than xCAT. This means the same infrastructure as code automation can be used to create baremetal and VM based platforms. Good practices and industry standard tools like Terraform and Ansible are being adopted to help make it easier to port these platforms both to other OpenStack clouds and non-OpenStack clouds.

Finally we look at some of the active development currently in flight, including work to add a temporal dimension to quota using OpenStack Blazar. The aim being to reduce overheads in rebalancing the allocations between multiple competing workloads.

Primary authors: GARBUTT, John (StackHPC); Mr PAUL, Browne (Cambridge University)

Presenters: GARBUTT, John (StackHPC); Mr PAUL, Browne (Cambridge University)

Session Classification: Cloud computing - Part 2