Contribution ID: **64**   Type: **Short talk**

Running Native HPC Applications on the Cloud

Wednesday 22 March 2023 11:20 (10 minutes)

In this project, we aim to enable Charm++ based HPC applications to run natively on a Kubernetes cloud platform. The Charm++ programming model provides a shrink/expand capability which matches well with the elastic cloud philosophy. We investigate how to enable running Charm++ applications with dynamic scaling of resources on Kubernetes. In order to run Charm++ applications on Kubernetes, we have implemented a Charm operator, very similar to Kubeflow's mpi-operator. The charm operator enables scaling of the number of pods in a job which isn't supported by the mpi operator since typically MPI applications do not support rescaling of resources at runtime. This operator also generates the nodelist in the correct format required by Charm++ programs for rescaling. The Charm++ application is launched in server mode to enable the injection of messages into the scheduler externally which is used to signal rescaling. The Charm operator handles allocation of resources and cleanup for all charm jobs on the Kubernetes cluster. For startup, it creates the launcher and worker pods for all jobs and performs monitoring for any change to a deployment configuration. We are implementing changes in the controller code which allow scaling of pods, i.e. shrinking or expanding the number of pods allocated to a Charm++ job. Currently, we have added support for making shrink/expand updates using the YAML file for the deployment. We use these shrink/expand updates to yaml script for testing our implementation. We are working on two modes for scaling, one where the pods are deleted on shrink and for expand new pods are created. In the second mode, we maintain a pool of worker pods where shrink releases worker pods to the pool of pods and these can be re-used for an expand request by another job in the context of the charm-operator.

JLESC topic

HPC+Cloud

Primary authors: Mr BHOSALE, Aditya (University of Illinois at Urbana-Champaign); Ms CHANDRASEKAR, Kavitha (University of Illinois at Urbana-Champaign); Dr BELLO-MALDONADO, Pedro (IBM); Dr COSTA, Carlos (IBM); Dr MISALE, Claudia (IBM); KOKKILA-SCHUMACHER, Sara (IBM); Prof. KALE, Laxmikant (University of Illinois at Urbana-Champaign); Dr KINDRATENKO, Volodymyr (University of Illinois at Urbana-Champaign)

Presenter: Dr KINDRATENKO, Volodymyr (University of Illinois at Urbana-Champaign)

Session Classification: Short Talks on Applications

Track Classification: Programming languages and runtimes