15th JLESC Workshop

Joint Laboratory for Extreme-Scale Computing

Contribution ID: **63**

Type: **Short talk**

Cloud-Bursting and Autoscaling for Python-Native Scientific and AI Workflows

Wednesday 22 March 2023 15:00 (10 minutes)

We have extended the Ray framework to enable automatic scaling of workloads on high-performance computing (HPC) clusters managed by SLURM© and bursting to a Cloud managed by Kubernetes®. Our implementation allows a single Python-based parallel workload to be run concurrently across an HPC cluster and a Cloud. The Python-level abstraction provided by our solution offers a transparent user experience, requiring minimal adoption of the Ray framework. Applications in Electronic Design Automation and Machine Learning are used to demonstrate the functionality of this solution in scaling the workload on an on-premises HPC system and automatically bursting to a public Cloud when running out of allocated HPC resources. The paper focuses on describing the initial implementation and demonstrating novel functionality of the proposed framework using three applications as well as identifying practical considerations and limitations for using Cloud bursting mode.

JLESC topic

HPC+Cloud

Primary authors: Mr LIU, Tingkai (University of Illinois at Urbana-Champaign); Dr ELLIS, Marquita (IBM); Dr COSTA, Carlos (IBM); Dr MISALE, Claudia (IBM); KINDRATENKO, Volodymyr (University of Illinois at Urbana-Champaign); Mrs KOKKILA-SCHUMACHER, Sara (IBM)

Presenter: Mr LIU, Tingkai (University of Illinois at Urbana-Champaign)

Session Classification: Short Talks on Distributed Resources

Track Classification: Programming languages and runtimes