

Cloud Bursting GPU workflow in support of Multi-Messenerg Astrophysics with IceCube

Friday, 28 August 2020 10:00 (30 minutes)

In Fall 2019, we performed the largest possible GPU burst across multiple commercial cloud providers, and all of their relevant regions. The goal of our NSF funded EAGER award (NSF OAC 1941481) was to achieve a scale of 80,000 V100 equivalent GPUs to process photon propagation simulations for the IceCube Neutrino Observatory for one hour, thus achieving fp32 Exaflop scale. Since then, we learned that it is rather unlikely that even the combination of the big three commercial cloud providers have that kind of on-demand capacity. In this talk, we will first report how we convinced ourselves that we have the capability to run the IceCube workflow at this scale across the cloud providers, including the necessary hundreds of terabytes of IO handling across the various relevant regions globally. After that, we present the actual scale achieved, and conclude with lessons learned from this exercise.

Primary author: Prof. WUERHWEIN, Frank (UCSD/SDSC)

Presenter: Prof. WUERHWEIN, Frank (UCSD/SDSC)

Session Classification: Infrastructure Clouds and Virtualisation Session

Track Classification: Infrastructure Clouds and Virtualisation