Birds of a Feather

SIGHPC - Big Data Chapter

Wednesday November 14th, 2018
12:15p - 1:15p
Room: D169

Session Leader: Stratos Efstathiadis, NYU
Additional Session Leader: Suzanne McIntosh, NYU

The **ACM SIGHPC Big Data Virtual Chapter** will host its ***third BoF*** at SC18. The BoF is open to everyone interested in the convergence of HPC and Big Data. We are pleased to announce the following presentations:

Andras Pataki, PhD
Senior Data Scientist, Flatiron Institute

Ceph implementation at Flatiron

Ceph is one of the newer distributed storage management systems that has seen very active development in recent years. Its open source development model together with a very flexible internal design that decouples the underlying reliable storage implementation (RADOS) from the various possible access methods provides a unique, high performance, reliable solution that may meet needs for users looking for a POSIX compliant file system, block devices for VM provisioning or simply object storage. After an overview of ceph, I will describe the ceph implementation at the Flatiron Institute, the in-house research arm of the Simons Foundation dedicated to the advancement of computational science across various disciplines.

Christopher N. Hill , PhD
Principal Research Engineer, Department of Earth, Atmosphere, and Planetary Sciences, MIT

Big Data and Big Models

A big part of future research is big data and big models, both solo and together. I will describe some collaborative activities at MIT, in Massachusetts and in the US Northeast region aimed at providing cyberinfrastructures for this future. The talk will highlight some research drivers and trends. It will also look at some practicalities of building a rounded, sustainable cyberinfrastructure ecosystem and community for a big future in big data and big models of all sorts.

Lucas A. Wilson, PhD
Dell/EMC HPC and AI Engineering

Getting Smarter Faster: The Intersection of HPC and AI

Deep learning is transforming many domains, from physical sciences to consumer businesses. It is a group of techniques that are heavily dependent on processing large quantities of data and computing differentiable functions – two places where HPC has excelled for decades. As the DL community looks toward HPC, what lessons can the DL community borrow to make creating intelligent systems faster? And as the HPC community looks toward DL, what lessons can the HPC community borrow to produce better scientific insights? Dell EMC's HPC and AI Engineering team has been working at the intersection of AI and HPC, looking at how best to run DL workloads on HPC systems and the challenges that bringing a new workload to an established model create.

Andy Watson
CTO of WekaIO

Storage Challenges for Machine Learning at Scale

We will discuss the storage challenges machine learning and Big Data analytics create once the data sets scale beyond what is possible to fit on a single GPU server. When companies are building out their machine learning strategy the focus tends to be on the model development and compute requirements. Storage tends to be an after-thought in the infrastructure design and often leads to poor GPU/CPU utilization. The talk will focus on what works and what pitfalls to avoid drawing from real-live customer use cases where data has to scale to tens of petabytes of training catalog.