Towards Understanding the Complex Correlation between ML and Systems

Murali Emani, Argonne National Laboratory

Abstract— **Machine learning (ML) is increasingly becoming critical in science applications. Emphasis on ML for systems and Systems for ML is highly stressed upon. There are emerging workloads with very interesting characteristics such as sparse data, AI-guided simulations. On the other hand, there is a surge of regular and AI-specialized hardware accelerators. There are unique capabilities along with challenges in leveraging these hardware. It is important to understand the deeper interactions across applications, middleware and hardware. Here, I will present our work on understanding how to better characterize Deep Learning workloads for scalability (Systems for ML) and auto-guided optimization of GPU Unified memories (ML for systems). I will also give an overview of our AI-hardware landscape.**

BIOGRAPHY

Murali Emani is a Computer Scientist in the Data Science group with the Argonne Leadership Computing Facility (ALCF) at Argonne National Laboratory. His research interests are in Scalable Machine Learning, Parallel programming models, High Performance Computing, Runtime Systems, Emerging HPC architectures, Online Adaptation. Prior, he was a Postdoctoral Research Staff Member at the Lawrence Livermore National Laboratory, US. He obtained his PhD from University of Edinburgh, UK. Murali published in top conferences including PACT, PLDI, CGO, SC and has three granted patents. He served as technical program committee member for conferences such as IPDPS, PACT, ICPP, CCGRID. He is a co-founder of MLPerf HPC working group, aimed to work with labs, academia and industry to create a benchmark suite for Scientific ML-based applications at scale. He was recently awarded DoE ASCR grant to develop a framework 'HPC-FAIR' to manage datasets and AI Models for Analyzing and Optimizing Scientific Applications.