

Machine Learning for I/O: Challenges and Opportunities

Friday, February 28, 2025 10:05 AM (25 minutes)

The increasing integration of machine learning and AI into HPC workflows presents both challenges and opportunities for I/O performance analysis. AI workloads, for example, generate I/O patterns that differ significantly from traditional HPC workloads, making it difficult to balance with the current I/O optimization configurations. On the other hand, machine learning also offers powerful tools to address challenges for predicting I/O performance, thus improving scheduling strategy, procurement, and application tuning.

In this talk, I will present two works related to machine learning for I/O. The first work is a benchmark extension to emulate AI workloads, and the second one utilizes a transfer learning workflow to create an effective I/O performance prediction with a fraction of the data and computing power compared to the predecessor works, which require resources not available to small and medium clusters.

Primary author: LIEM, Radita (RWTH Aachen University)

Presenter: LIEM, Radita (RWTH Aachen University)