

Title: ParaLog: Transparent Acceleration of Scientific Applications

Abstract: Scientific applications are output-intensive. During execution, they alternate between two blocking phases, performing computation and outputting results. As such, they are sensitive to low storage throughput because a slow output leads to idle time in computation. While these applications traditionally run on High-Performance Computing (HPC) systems with shared, high network and storage bandwidth, these assumptions are not necessarily true in modern setups, such as a cloud-HPC environment. In this talk, we discuss how to exploit the characteristics of scientific applications to accelerate their output phases using coordinated logging. This creates a new design point in a distributed logging system that does not compromise data safety and deployability. We also look at how this method can extend to support machine learning applications that require frequent checkpointing.