

I/O Performance Analysis Framework on Measurement Data from Scientific Clusters

Michelle Koo
University of California, Berkeley
Lawrence Berkeley National Laboratory

ABSTRACT

Due to increasing scales in data volumes, number of machines, and complex workflow in science applications, it has become more challenging to diagnose job performance and system performance on scientific clusters. This project is motivated by the observations that I/O performance analyses can be conducted from monitored performance measurement data from scientific clusters. Studies of I/O performance behavior have been conducted on the Palomar Transient Factory (PTF) application, an automated wide-field survey of the sky to detect variable and transient objects, by analyzing measurement data collected on NERSC Edison. Visualization tools were developed to aid in identifying I/O performance bottlenecks in the PTF data analysis pipeline. This work led to building an interactive I/O performance analysis framework for measurement data from scientific clusters to further identify performance characteristics and bottlenecks in scientific applications.