

Abstract

Efficient utilization of HPC clusters is crucial due to their high upfront costs and on-going operational expenses. On the CLAIX systems at the RWTH Aachen University, performance monitoring data is collected to analyze hardware utilization for this reason. By predicting the power consumption and heat dissipation of the cluster with the monitoring data, the cooling controls of the cluster could be optimized and automated. To achieve this goal, it is first necessary to verify the reliability and correctness of the data. This work aims to fill the gap by analyzing the reliability of the time series data by searching for missing values. The correctness of the data is assessed by searching for implausible values. This work also presents preprocessing methods to repair missing and implausible values, as well as an analysis of the correlations between performance-related and power-related metrics which could be valuable for future predicting modeling approaches.