Fair Resource Allocation for Data-Interval Computing in the Cloud

To address the computing challenge of 'big data', a number of data-intensive computing frameworks (e.g., Map Reduce, Dryad, Storm and Spark) have emerged and become popular. YARN is a de facto resource management platform that enables these frameworks running together in a shared system. However, we observe that, in cloud computing environment, the fair resource allocation policy implemented in YARN is not suitable because of its memory less resource allocation fashion leading to violations of a number of good properties in shared computing systems. This paper attempts to address these problems for YARN. Both single level and hierarchical resource allocations are considered. For single-level resource allocation, we propose a novel fair resource allocation mechanism called Long-Term Resource Fairness (LTRF) for such computing. For hierarchical resource allocation, we propose Hierarchical Long-Term Resource Fairness (H-LTRF) by extending LTRF. We show that both LTRF and H-LTRF can address these fairness problems of current resource allocation policy and are thus suitable for cloud computing. Finally, we have developed LTYARN by implementing LTRF and H-LTRF in YARN, and our experiments show that it leads to a better resource fairness than existing fair schedulers of YARN.