Optimal workload allocation model for scheduling divisible data grid applications

ABSTRACT

In many data grid applications, data can be decomposed into multiple independent sub-datasets and distributed for parallel execution and analysis. This property has been successfully employed using Divisible Load Theory (DLT), which has been proved a powerful tool for modeling divisible load problems in data-intensive grids. There are some scheduling models that have been studied but no optimal solution has been reached due to the heterogeneity of the grids. This paper proposes a new model called the Iterative DLT (IDLT) for scheduling divisible data grid applications. Recursive numerical closed form solutions are derived to find the optimal workload assigned to the processing nodes. Experimental results show that the proposed IDLT model leads to a better solution than other models (almost optimal) in terms of makespan.

Keyword: Scheduling; Divisible load theory; Data grid