Title: Makespan Reduction for Dynamic Workloads in Cluster-based Data Grids using Reinforcement-Learning Based Scheduling

Authors: Mahshid Helali Moghadam, Seyed Morteza Babamir

PII: S1877-7503(17)30231-4
DOI: https://doi.org/10.1016/j.jocs.2017.09.016
Reference: JOCS 769

To appear in:

Received date: 28-2-2017
Revised date: 22-9-2017
Accepted date: 24-9-2017

Please cite this article as: Mahshid Helali Moghadam, Seyed Morteza Babamir, Makespan Reduction for Dynamic Workloads in Cluster-based Data Grids using Reinforcement-Learning Based Scheduling, Journal of Computational Science https://doi.org/10.1016/j.jocs.2017.09.016

Makespan Reduction for Dynamic Workloads in Cluster-based Data Grids using Reinforcement-Learning Based Scheduling

Mahshid Helali Moghadam and Seyed Morteza Babamir*

Department of Computer Engineering, University of Kashan, Kashan, Iran

Corresponding author: *babamir@kashanu.ac.ir

Highlights

- In this study, a two-step adaptive task scheduling framework based on data awareness and reinforcement learning is proposed for cluster-based data grids
- According to simulation results, the proposed adaptive scheduling strategy gives better performance with increasing the number of submitted tasks, in comparison with other common scheduling strategies for different replication strategies
- According to the first simulation experiment, the proposed two-step adaptive scheduling stratgey has the lowest mean response time for various number of tasks with LRU replication
- With increasing the number of submitted tasks, the amount of performence improvement will be more.
- The proposed adaptive scheduling with LFU replication has also better performance compared with other scheduling strategies and outperforms the other ones significantly for large number of submitted tasks
- The proposed adaptive scheduling has higher perfomance with the second exploration policy for high number of submitted tasks as cen be seen in real application environments
- Using learning-based policy increases the adaptation of the proposed approach to dynamic nature of the environment and high variety of submitted tasks.

Abstract. Scheduling is one of the important problems within the scope of control and management in grid and cloud-based systems. Data grid still as a primary solution to process data-intensive tasks, deals with managing large amounts of distributed data in multiple nodes. In this paper, a two-phase learning-based scheduling algorithm is proposed for data-intensive tasks scheduling in cluster-based data grids. In the proposed scheduling algorithm, a hierarchical multi agent system, consisting of one global broker agent and several local agents, is applied to scheduling procedure in the cluster-based data grids. At the first step of the proposed scheduling algorithm, the global broker agent selects the cluster with the minimum data cost based on the data communication cost measure, then an adaptive policy based on Q-learning is used by the local agent of the selected cluster to schedule the task to the proper node of the cluster. The impacts of three action selection strategies have been investigated in the proposed scheduling algorithm, and the performance of different versions of the scheduling algorithm regarding different action selection strategies, has been evaluated under three types of workloads with heterogeneous tasks. Experimental results show that for dynamic workloads with varying task submission patterns, the proposed learning-based scheduling algorithm gives better performance compared to four common scheduling algorithm, Queue Length (Shortest Queue), Access Cost, Queue Access Cost (QAC) and HCS, which use regular combinations of primary parameters such as, data communication cost and queue length. Applying a learning-based strategy provides the scheduling algorithm with more adaptability to the changing conditions in the environment.

Key words: Data grid, Data-intensive task scheduling algorithm, Data communication cost, reinforcement learning.