

Staging-Based Data Management for Extreme Scale Coupled Scientific Workflows

Qian Sun
Dept. of Computer Science

4/5/2017 at 02:00 pm
CoRE A (301)

Abstract

Advanced scientific workflows running at extreme scale on high end computing platforms are providing new capabilities and new opportunities for insights in a wide range of application domain. These workflows compose multiple simulation, data analysis and other application components that require data sharing and exchange at runtime. However, due to the increasing data volumes and associated I/O costs, the traditional disk I/O-base approach for data sharing between the components of these workflows is becoming infeasible. Recently, in memory data staging has emerged as an attractive approach to address these challenges. The increasing systems scales, the overall complexity of current scientific workflows, as well as the complexity of the coupling and coordination behaviors between workflow components are presenting several new data management challenges that are impacting the effectiveness and efficiency of staging-base solutions. This thesis addresses in-staging data management challenges to enable effective and efficient staging-based scientific workflows, including (1) query-driven analytics workflows (2) workflows with dynamic runtime behaviors (3) asynchronous coupling of task-based scientific workflows. The research components have been prototyped and experimentally evaluated using real application workflows on current high end computing systems, and their effectiveness and performance is demonstrated.

Defense Committee: Prof. Manish Parashar(chair), Prof. Ulrich Kremer and Prof. Zheng Zhang