Fault-tolerant and data-intensive resource scheduling and management for scientific applications in cloud computing

ABSTRACT

Cloud computing is a fully fledged, matured and flexible computing paradigm that provides services to scientific and business applications in a subscription-based environment. Scientific applications such as Montage and CyberShake are organized scientific workflows with data and compute-intensive tasks and also have some special characteristics. These characteristics include the tasks of scientific workflows that are executed in terms of integration, disintegration, pipeline, and parallelism, and thus require special attention to task management and data-oriented resource scheduling and management. The tasks executed during pipeline are considered as bottleneck executions, the failure of which result in the wholly futile execution, which requires a fault-tolerant-aware execution. The tasks executed during parallelism require similar instances of cloud resources, and thus, cluster-based execution may upgrade the system performance in terms of make-span and execution cost. Therefore, this research work presents a cluster-based, fault-tolerant and data-intensive (CFD) scheduling for scientific applications in cloud environments. The CFD strategy addresses the data intensiveness of tasks of scientific workflows with cluster-based, fault-tolerant mechanisms. The Montage scientific workflow is considered as a simulation and the results of the CFD strategy were compared with three well-known heuristic scheduling policies: (a) MCT, (b) Max-min, and (c) Min-min. The simulation results showed that the CFD strategy reduced the make-span by 14.28%, 20.37%, and 11.77%, respectively, as compared with the existing three policies. Similarly, the CFD reduces the execution cost by 1.27%, 5.3%, and 2.21%, respectively, as compared with the existing three policies. In case of the CFD strategy, the SLA is not violated with regard to time and cost constraints, whereas it is violated by the existing policies numerous times.

Keyword: Scientific workflows; Scheduling; Fault-tolerant; Montage; Clustering