ETSCC015 An Efficient and Scalable Framework for Processing Remotely Sensed Big Data in Cloud Computing Environments

Abstract

The large amount of data produced by satellites and airborne remote sensing instruments has posed important challenges to efficient and scalable processing of remotely sensed data in the context of various applications. In this paper, we propose a new big data framework for processing massive amounts of remote sensing images on cloud computing platforms. In addition to taking advantage of the parallel processing abilities of cloud computing to cope with large-scale remote sensing data, this framework incorporates task scheduling strategy to further exploit the parallelism during the distributed processing stage. Using a computation- and data-intensive pan-sharpening method as a study case, the proposed approach starts by profiling a remote sensing application and characterizing it into a directed acyclic graph (DAG). With the obtained DAG representing the application, we further develop an optimization framework that incorporates the distributed computing mechanism and task scheduling strategy to minimize the total execution time. By determining an optimized solution of task partitioning and task assignments, high utilization of cloud computing resources and accordingly a significant speedup can be achieved for remote sensing data processing. Experimental results demonstrate that the proposed framework achieves promising results in terms of execution time as compared with the traditional (serial) processing approach. Our results also show that the proposed approach is scalable with regard to the increasing scale of remote sensing data.