International Conference on Advances in Materials, Computing and Communication Technologies
In Association with International Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Big Data Equilibrium Scheduling Strategy in Cloud Computing Environment

S. L. Jenishaliphni[1], Mrs. Anchana B.S.[2]

[1]Department of Computer Science and Engineering, Marthandam College of Engineering and Technology, Kuttakuzhi, Tamil Nadu, India

[2]Professor and Head, Department of Computer Science and Engineering, Marthandam College of Engineering and Technology, Kuttakuzhi, Tamil Nadu, India

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.

Keywords : Directed Acyclic Graph

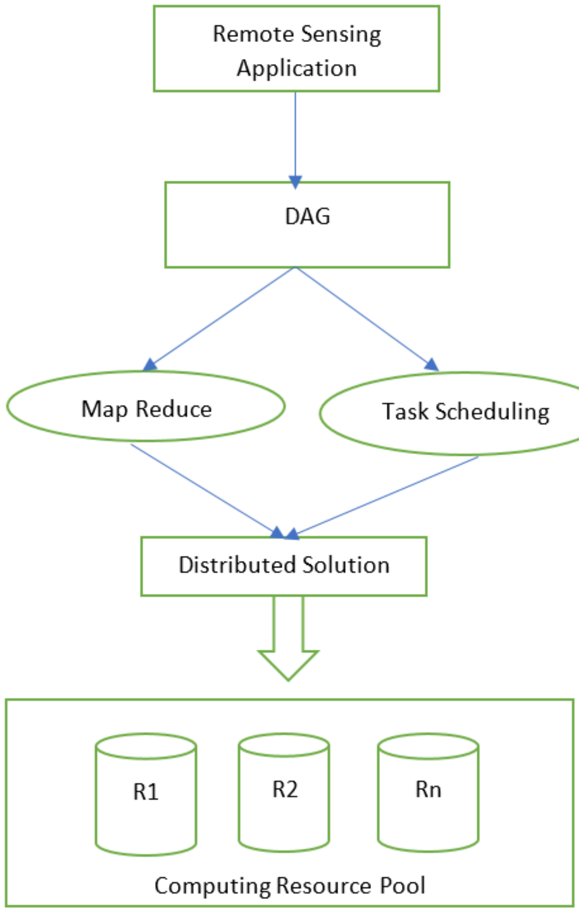
I. INTRODUCTION

The amount of data available from remote sensing systems is increasing at an extremely fast pace due to recent advances in modern earth observation technologies. The tremendous amount of remotely sensed data has posed serious challenges for efficient and scalable processing of remotely sensed data in the context of various applications. Considering the high volume and fast generation velocity of remotely sensed data, the processing of such data can be naturally regarded as a big data problem.

Cloud computing offers the capability to tackle big data processing by means of its distributed computing mechanism. The use of cloud computing for the analysis of large-scale remote sensing data repositories represents a natural solution, as well as an evolution of previously developed techniques. With the continuously increasing demand for massive data processing in remote sensing applications, there have been several efforts in the literature oriented toward exploiting cloud computing infrastructure for processing remote sensing big data. For example, Wu et al. proposed a parallel and distributed implementation of the principal component analysis algorithm for hyperspectral dimensionality reduction. Based on the MapReduce parallel model, this implementation utilizes Hadoop's distributed file system (HDFS) to realize distributed storage and uses Apache Spark to perform parallel computing of hyperspectral data. Quirita et al. [9] proposed a new distributed architecture for supervised classification of large volumes of earth observation data. This architecture supports distributed execution, network communication, and fault tolerance to the user. However, the existing cloud computing solutions are mainly developed toward solving big data problems in a specific category of remote sensing applications.

In cloud computing environment, big data transmission channel equalization scheduling technology is taken based on big data transmission channel multi-mode control and symbol modulation and demodulation technology. In cloud computing environment, big data transmission channel is affected by electromagnetic medium interference and multi-path effect of channel output[3]. It is easy to produce inter-symbol interference, which leads to communication channel imbalance. It is necessary to optimize the design of big data equalization scheduling. The main methods of big data communication scheduling in cloud computing environment are frequency domain decomposition scheduling method, load balancing scheduling method, Neural network scheduling and so on [, through the design of intelligent scheduling method of big data and the equalization design of big data communication channel in cloud computing environment, the equalization scheduler is constructed to realize cloud computing processing of massive data. Some scholars have studied the design of equalization scheduler, and some achievements have been made. Wherein, in reference [4], a data classification scheduling model is proposed based on naive Bayes in cloud computing environment, which optimizes the big data transport scheduling network model and designs channel equalization under cloud computing, which has better scheduling performance.

II. ARCHITECTURAL DESIGN



III. RELATED WORK

This section introduces the concept of task scheduling in distributed processing of a remote sensing application. A scheduling model consists of an application, a target computing environment, and evaluation criteria for scheduling. In general, an application is fundamentally composed of sets of tasks, e.g., the fusion process of the DNN-based pan sharpening method in Fig. 1. Tasks within a particular flow may have dependencies among them, while tasks from different flows tend to have very few or no dependencies among them. In this manner, an application is usually represented by a DAG, denoted by G =(V, E), where V ={ v1,v2,...,vn} is a set of tasks to be executed, and E ={ (i, j)}is a set of edges between the tasks. Each edge (i, j)∈ E specifies a precedence constraint on two connected tasks: task vi must be completed before vj can start. In addition, each task is associated with a weight value Di representing the duration time of task vi. Fig. 5 provides a simple example of a task graph. Let Si and Sj denote the starting times for vi and vj, respectively. The precedence relationship exerts the following set of constraints:

The scheduling problem is fundamental to determine the mapping function that maps each task to a computing resource for minimizing the scheduling goal. A schedule of a DAG on available computing resources can be illustrated as a Gantt chart that consists of a list of all computing resources, and for each computing resource, a list of assigned tasks of which the execution order is determined by task starting and ending times [35]. In this paper, the scheduling goal is to minimize the total completion time of a remote sensing application. This performance measure is known as the make span of application, which is determined by the maximum ending time of any task.

IV. METHODOLOGY

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.

1) This paper provides a scheduling-enabled cloud computing solution to the efficient processing of remote sensing applications in which the workflows can be well represented by DAGs.
2) It develops an optimization framework that searches for an optimal distributed processing solution, including the numbers of partitions for partitionable tasks and the assignments of workers executing all tasks, under the constraint of limited computing resources.
3) It further develops a cost-effective, QEA-based metaheuristic scheduling algorithm to solve the above-mentioned optimization problem and obtain the optimized distributed processing solution.
4) Experimental results demonstrate that the proposed framework achieves promising results in terms of execution time and speedups as compared with the serial processing approach and is scalable to the increasing scale of remotely sensed data sets.

V. CONCLUSION

This paper presents a new cloud computing framework that integrates a distributed processing mechanism and a task scheduling strategy into an optimization procedure to enable efficient and scalable processing of large-scale remotely sensed data. As a case study, we optimize a DNN-based fusion method on Spark to verify the efficiency of the proposed framework. Experimental results demonstrate that the proposed framework achieves promising speedups as compared with the serial processing approach. More importantly, this framework is also scalable with regard to the increasing scale and dimensionality of remote sensing data.

VI. REFERENCES

[1]. KHABBAZIBASMENJ A, HASSANIEN A, VOROBYOV S, et al. Efficient transmit beamspace design for search-free based DOA estimation in MIMO radar[J]. IEEE Transactions on Signal Processing, 2014, 62(6): 1490-1500
[2]. SAMMARTINO P F, BAKER C J, and GRIFFITHS H D. Frequency diverse MIMO techniques for radar[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(1): 201-222.
[3]. WANG W Q and SHAO H. Range-angle localization of targets by a double-pulse frequency diverse array radar[J]. IEEE Journal of Selected Topics in Signal Processing, 2014, 8(1): 106-114.
[4]. WANG W Q and SO H C. Transmit subaperturing for range and angle estimation in frequency diverse array radar[J]. IEEE Transactions on Signal Processing, 2014, 62(8): 2000-2011.
[5]. WANG Youcheng, ZHANG Feng, FANG Guangyou, et al. A novel ultrawideband exponentially tapered slot antenna of combined electric-magnetic type[J]. IEEE Antennas and Wireless Propagation Letters, 2016, 15: 1226-1229.