# A New Framework of Cluster-Based Parallel Processing System for High-performance Geo-computing

Yan Ma[1] Dingsheng Liu[1] Jingshan Li[1]

[1]（Center for Earth Observation and Digital Earth，Chinese Academy of Sciences, 100086, Beijing）

{yanma, dsliu, jsli}@ceode.ac.cn

## Abstract

Up to now, it still remains a big challenge for us to build a high performance geo-computing system which should be of high processing performance and be also easily used by normal users who may have no knowledge about high performance system. The main problems for building such system is that the geo-computing always have to face the huge data, various algorithms with high complexity, geo-information closely related to data and processing procedure, as well as the high relevancy between data and algorithms. All of these problems make the system constructing more difficult in building a general procedure for most processing task and making such procedures be easily understand for normal users.

In this article, based on the analysis of the characteristics of Geo-computing and the current feasible technical approaches to general high performance computing, a new architecture for the high-performance geo-computing is illustrated to demonstrate our solutions to the above key issues.

To properly settle these main issues above, several technical details should be deeply probed into, including how to eliminate the I/O performance bottleneck caused by handling massive large-scale image data related to data intensive algorithms, how to deal with the data managing problem result from the relevancy between geo-information and remote sensing image data, and how to build a serious parallel processing models in order to get high parallel efficiency for a variety of algorithms and also make the parallelization transparent to users.

Based on above researches, a High Performance Geo-data Object Storage System (HPGOSS) base on parallel file System is presented. According to the different repetitive computing mode, and different relationship between data distribution and computing, two parallel programming models for fast parallelization of geo-computing algorithms are extracted. These models also include optimization on multi-core processor and performance balance between I/O and CPU. In addition, the job scheduling strategy and workflow engine are also discussed.

Finally, a new system framework for high performance geo-computing is constructed with the combination of these key techniques. And such system could provide a parallel geo-computing environment with high performance, easy to use, optimal resource utilization, and high scalability.

**Key Words:** high-performance computing; geo-computing; remote sensing image processing; system framework; parallel file system; parallel programming model