

# High Performance Data-Intensive Remote Sensing Image Processing

Lizhe Wang, CEODE/CAS

Recently, with the proliferation of data, Remote Sensing (RS) data processing becomes extremely challenging because of the massive big image data. In particular, many time-critical applications like disaster monitoring even require real-time or near real-time processing capabilities. Remote Sensing (RS) data processing is recognized as a typical data-intensive application. However, parallel cluster systems are characterized by increasing scale and multilevel parallel hierarchy. Parallel processing of data-intensive applications like massive remote sensing image processing on parallel systems is bound to be especially difficult and challenging. The dependencies between the computation and associated RS data in RS algorithms, together with the concerning for the detailed hierarchical architecture of parallel systems (message passing model for inter-nodes and the shared memory model form intra-node) will certainly lead to poor programmability of parallel RS algorithms. Also, it is no longer possible to load the entire large RS dataset into the limited memory of single node, and the data communication across nodes of big RS datasets with complex data structures are not efficiently supported by current MPI runtime system. Furthermore, the intensive large amount of data throughput in short period of time will pose large I/O challenge for parallel system. The complex data accessing mode of algorithms, together with the poor data locality will make the data accessing even more difficult and inefficient. Recently, the related researches mainly focus on only the parallelization of some specific RS algorithms on parallel systems with complex hierarchical architecture. But the parallel runtime systems especially for data-intensive remote sensing image processing which dealing with the problems of programmability, data load in and communication of big RS data, and the complex data accessing mode are rare involved.

To deal with the above issues induced by the data-intensive characteristic of remote sensing image processing, we are probing a new high performance parallel processing platform especially for remote sensing image processing. This runtime system provides a parallel file system for remote sensing image with multiple data slicing and copy strategy. This file system will offer complex data accessing mode for RS algorithms, data locality and finally high throughput of data I/O. Also, we provide a generic parallel programming for massive RS data processing in this system to offer elegant parallel programmability of RS algorithms. Developers without extensive parallel computing technologies can write efficient parallel RS programs on this platform without concerning for parallel computing details. In addition, we propose a distributed RS data model for managing and communication of large RS datasets with complex data structure. The distributed RS data could provide a global view of whole data whose sliced data blocks are scattered among nodes. Moreover, by data serialization and RMA (Remote Memory Access), the data templates could also offer a simple and effective way to distribute and communicate massive remote sensing data with complex data structures.