A HIGH PERFORMANCE REMOTE SENSING RETRIEVAL APPLICATION ON AN INSTITUTIONAL DESKTOP GRID

Wei Wan[1], *Yong Xue*[2,3], *Senior Member, IEEE, Jianning Guo*[1], *Shuning Lu*[1], *Jie Guang*[2,3]

[1]China Center for Resource Satellite Data and Application, Beijing, China
[2]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Institute of Remote Sensing Applications, Chinese Academy of Sciences, PO Box 9718, Beijing 100101, China
[3]Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
{Emails: eric_104@163.com, yxue@irsa.ac.cn}

Quantitative retrieval of aerosol property from remotely sensed data is a both data and computing intensive scientific application, where the complexities of processing, modeling and analyzing large volumes of remotely sensed data sets have significantly increased computation and data demands.

The remote sensing quantitative retrieval within the atmospheric aerosol domain is an example of the type of scientific application that will benefit from an advanced Grid computing (Foster et al., 2001). Grid computing takes advantage of many networked computers to model a virtual computer architecture that is able to distribute process execution across a distributed infrastructure. Grids gather the computational resources of separate computers connected by a network to solve large-scale computation problems. Grids provide the ability to perform computations on large data sets by breaking them down into many smaller ones.

Institutional desktop Grid comprised by the desktop machines of an institution (academic or corporate) can act as an important source of computing power for local user, if the high volatility of such resources is properly harnessed. Office PCs sit unused during idle hours, which can easily count for large portion of computing time. An institutional desktop Grid using existing office PCs has the potential to be an enormous computing resource, particularly considering current and planned connections of network.

In this paper, we focus on making desktop Grids adapted to remote sensing application. For that purpose, a middleware framework for remote sensing retrieval named the Remote Sensing Information service grid Node (RSIN) is described. We proposed the design of the Remote Sensing Information service Grid Node (RSIN) framework (Xue et al., 2008) and the High Performance Aerosol property Retrieval Software (HiPARS) application (Wan, 2008). RSIN framework meets the needs of the remote sensing community by providing high performance, ease of use, and extensibility.

This framework has features:

1) Run on high throughput computing Grid,
2) Enable a workflow system for task management and data placement,
3) Enable accompanying unified data-and-computation-schedule algorithm and a task partition algorithm, which help load balancing between and within workflow steps,
4) Be usable without advanced knowledge of Grid architecture or parallel algorithms,
5) Provide a mechanism for code reuse and module Grid-enabling, and
6) Provide a graphical environment and tools that facilitate fast model development, data visualisation and performance analysis.

In this paper, we discussed RSIN framework aimed at addressing challenges that the remote sensing community faces. To solve data-intensive issue, RSIN framework views data placement tasks and computational jobs as equally important, a workflow manager executes data placement, distributing data so that it stays separate and asynchronous from computational execution. A unified Data and Computation Scheduling algorithm is proposed for load balancing between and within workflow levels. A Workload Estimation and Task Partition algorithm, combining static planning and dynamic scheduling, is devise to realize the parallelism of generic remote sensing algorithms on a Grid platform.

The initial efforts were made to build a remote sensing retrieval application on the RSIN framework for dealing with climate change, named High Performance Aerosol property Retrieval Software (HiPARS). HiPARS uses integration of applications that access the backend of desktop Grid comprising of office PCs with a client. The HiPARS remote sensing application was built on RSIN framework, which uses the loose-coupled architecture of a desktop client and high throughput computing (HTC) (Basney et al., 1997) and Grid backend. It allows for fast development by enabling existing code and new algorithms, and provides a familiar graphical environment for remote sensing users. A graphical environment is provided which allows users to specify a workflow for the application by creating and connecting pluggable modules, launch jobs, select parameters and data, analyze performance and visualize results. Issues dealing with how to achieve the parallelism, schedule tasks, and pass data are hidden from the user. With HiPARS, advanced understanding of Grid system has been strayed away from remote sensing users.

The proposed solution is to transition the current PCs to a desktop Grid that can be cost-effectively sustained. It will be accessible for processing satellite data in quasi real-time, for parameter retrieving or for parallel processing efforts.

Key words: Desktop Grid; Grid computing; Quantitative retrieval; Aerosol optical thickness; Remote Sensing