Study on Grid Enabled Aerosol Quantitative Remote Sensing Retrieval System

Wei Wan\textsuperscript{a,c}, Yong Xue\textsuperscript{a,b,*}, Yingjie Li\textsuperscript{c}, Ying Wang\textsuperscript{b,c}, Jianwen Ai\textsuperscript{a,c}, Jie Guang\textsuperscript{b,c}, and Linyian Bai\textsuperscript{a,c}

\textsuperscript{a}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, P.O. Box 9718, Beijing 100101, China

\textsuperscript{b}Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N78 DB, UK

\textsuperscript{c}Graduate School of the Chinese Academy of Sciences, Beijing, China

Email: \{y.xue@londonmet.ac.uk, eric_104@163.com\}

Quantitative retrieval of aerosol property from remotely sensed data is a data-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. While Grid has been a prominent technique to tackle computational issue, little work has been done on making grid computing adapted to remote sensing application. In this paper, a middleware framework for remote sensing retrieval named the Remote Sensing Information Service Grid Node (RSIN) is described. RSIN is targeted at producing applications for high throughput computing Grid. It enables a workflow management system for data placement. The accompanying unified data-and-computation-schedule algorithm helps load balancing between and within workflow steps (Xue et al., 2008). A workload estimation and task partition algorithm executes generic remote sensing algorithm in parallel over partitioned datasets. RSIN framework which meets the needs of the remote sensing community by providing high performance, ease of use, and extensibility. This framework has functionalities:

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.

The initial efforts are 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) (Wan, 2008). HiPARS uses integration of applications that access the Grid backend with a desktop client. 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. HiPARS customizes the RSIN middleware to the remote sensing domain by changing the terminology presented in the interface to be more familiar to the remote sensing user. It provides basic functions for selecting datasets through file lists, choosing a region of interest, calibrating data, geo-referencing, stitching mosaic images, enhancing image, screening cloud, converting map projection, and visualization of results. Additionally, a Grid Service Interface assists in development of a grid application by providing a place for describing the dataflow graph, launching jobs on the parallel computer, visualizing results in realtime, and analysing performance after job completion. HiPARS is built upon three major technologies: the desktop based GUI client, the Job Management Framework (JMF) and the D-W-S (Decompose, Wrap and Submit) method. JMF is a low-level middleware for managing tasks created by user, maintaining task lists, and mapping them to job queues on parallel computers. D-W-S is a routine used to enable existed applications or legacy codes on the grid environment.

HiPARS GUI client three basic functions:
1) User Interface - The UI not only responds to user’s interaction, but also creates views for satellite data, map and spatially explicit image results.
2) Task Launcher - The launcher is a user-event handler, which creates tasks that are invoked by GUI menu inputs, and sends them into Job Manage Framework.

3) Graph Editor - The editor is a tool for graphically describing the remote sensing quantitative retrieval workflow.

A case study shows that significant improvement of system performance can be achieved with our implement. It also gives a perspective on the potential of applying Grid computing practices to remote sensing problems.

**Keywords:** Data intensive; Grid; High Throughput; Remote Sensing; Quantitative Retrieval; Aerosol

**Acknowledgments:**

This work was supported in part by the NSFC under grants no. 40671142 and 40471091, by the CAS under grant no. KZCX2-YW-313, and by the MOST China under grant number (2007CB714407).