THE DESIGN OF DYNAMIC GRID WORKFLOW WEB PORTAL FOR REMOTE SENSING INFORMATION SERVICE

Jianwen Ai¹,2,4, Yong Xue¹,3*, Wei Wan¹,4, Guoping Lei², Shutao Hou², Yuqing Li², Fenghai Yang², and Zhihua Wang²

¹State Key Laboratory of Remote Sensing Sciences, 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
²College of Resources and Environmental Sciences, Northeast Agricultural University, Harbin, 150030, China
³Department of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
⁴Graduate University of Chinese Academy of Sciences, Beijing 100049, China

{E-mail: neau_ajw@hotmail.com; y.xue@londonmet.ac.uk}

ABSTRACT

Remote Sensing data processing and analysis involve multiple algorithm process procedures. Traditional Remote Sensing algorithms are tight-coupling. It is possible for the Grid Workflow to integrate distributed data and algorithms among idle computational resources in a Grid computing environment. This paper firstly introduces the workflow description language and the features of remote sensing algorithms. Then, it gives the status of workflow research and analyses the related research of Grid workflow applied to remote sensing domain. It presents the necessity to implement an GUI of dynamic Grid Workflow web portal applied to Remote Sensing information service. During the design, the authors use an 8-tuples to illustrate the relationships of the different parts of such a workflow. Finally, the authors implements the GUI of Grid workflow described.

Index Terms—Workflow, Grid, Remote Sensing, Geocomputing, Web portal

1. INTRODUCTION

Remote Sensing data processing and analysis are the indispensable part of the Geocomputing, and involve multiple algorithm processes. The high diversity of the algorithms and the high costs of high-performance supercomputers drive us to hunt for share of algorithms and computing systems. In order to utilize the existing algorithms, we require a method which can transfer the tight-coupling and system-centric algorithms into loose-coupling algorithms, a service which access the existing algorithms from Internet, a communication of among algorithms running on different platforms and written in different development language, and a tool which can use interactive graphical editors to present the executing relationships of algorithms on human-friendly diagrams. Fortunately, workflow technologies make it possible. Using workflow technology, we can construct a remote sensing information processing environment to integrate the distributed data and computational resources.

Today, Workflow Management Coalition gives many standards for business workflow being used. The workflow executed in distributed Grid infrastructures provides a technology to “glue” distributed data, heterogeneous computing resources, and users from many research institutions together. The workflow description language has been well formatted and standardized by the standard organizations, such as WfMC’s WPDL/XPDL, BPMI’s BPML, IBM’s WSFL, ebXML’s BPSS, RosettaNet’s PIP, BPEL4WS, GSWL, and so on[1,2]. Then it is very hard to
draw up a single standard, because almost all workflow and business process management systems have their own modeling methodology and system[1,2]. These languages force a formal description of semantics of a given language, which makes us difficult to grasp them without the technical expertise.

Remote sensing data and information are significantly different from those in other disciplines. Remote sensing data are different from other data because of the distributed, multi-spectral, low-correlated and heterogeneous characters. It has a set of standards specifically for handling data and information. The process of remote sensing data is complicated and time consuming[3]. Remote sensing software packages such as ARCGIS, ENVI, and ERDAS are expensive, and the use of them needs special skills[3]. On the other hand, researchers might prefer to pay for only the function modules which he or she is interested in, or have a third party to do the pre-processing steps or offer some computing power on demand[3]. Thus we need a system that provides individual function and computing power, allow user to pay for only certain functions[3]. Besides, the business workflow is dominated by control, events, and task-oriented approaches, making them less suitable for the modeling challenges of remote sensing Grid workflows. Existing workflow tools are not designed for remote sensing data processing and analysis. All of these restrict the applicability of existing workflow tools in remote sensing domain.

2. STATUS OF WORKFLOW RESEARCH

There are many scientific projects related to Grid workflows, (http://www.gridworkflow.org/snips/gridworkflow/space/Projects accessed in June 2008):

1) EU FP5 projects: including GRASP, GRACE, GRIA, GEMSS, OpenMoGIRD, GridLab
2) EU FP6 projects: including A-WARE, Akogrimo, CoreGRID, DEISA, Diligent, EGEE, ELeGI, HPC4U, InteliGrid, K-WfGrid, NextGrid, OntoGrid, Provenance, SIMDAT, UniGrids, SECSE, SODIUM, TrustCom
3) National Projects: including gridbus (Australia), Askalon (Austria), GRAAL, ProActive, and YML (France), C3Grid, D-Grid, DGI, EnterpriseGrids, Fraunhofer Resource Grid, Instant-Grid and MediGrid (Germany), Grid.it, INFNGrid, Knowledge Grid (Italy), JOpera (Switzerland), myGrid-Taverna, myGrid-WOSE/Triana (UK), GriPhyN/CAT (US), and R-Net (China)
4) Cross-Project Collaborations: including Kepler

By analyzing these Grid workflow projects, we can classify the Grid workflow research according to the implementation: 1) consists of a static Grid workflow. We must write workflow codes. The workflow cannot be modified during the execution; 2) consists of several parallel workflows or nesting workflows, which are still a static Grid workflow. We must still write workflow codes; 3) is Generated by dynamic tools, which can provide service according to users needs. Some scientific projects, like Kepler, can generate a new workflow about scientific computing with user’s dragging the mouse. But the users must setup some software and cannot generate a workflow by Internet explore with user’s dragging the mouse.

About the related research of specialized Grid workflow applied to remote sensing data processing and analysis, Abramson et al. implemented an atmospheric sciences workflow with Web Services by overloading the normal file IO operations to allow them to work in the Grid[4]. One of the more significant achievements of the experiment is that they managed to do this without any changes in the source codes of the models[4]. The Geosciences Network (GEON) project implements a LiDAR Workflow. The GEON LiDAR Workflow (GLW) provides a window to a user-defined selection of raw data that may be modeled into unique, user-defined DEMs or visualizations[5]. The GLW can do this by democratizing the data though multiple super computers[5]. These two workflow project are static workflow. Users can not order their function by GUI in Internet browser.

3. DESIGN OF WEB PORTAL

In order to solve these problems and access the distributed and heterogeneous resources and services, designing a GUI of dynamic Grid workflow web portal for remote sensing information service is necessary. For many users are not only short of the technical expertise to use the existing Grid workflow components, but also have not specific knowledge of the remotely sensed data processing methods. Our remote sensing information analysis and service Grid node (RSIN) have solved the remote cooperation, resource sharing, and management physically distributed. Our ongoing research is to provide the users an intuitive graphical user interface (GUI) to express steps of tasks being processed via Internet, just like the “order form” of tasks. Such a design of Workflow is necessity. It not only provides users a new service shared with algorithms and distributed resources by an easy-to-use interfaces, but also gives a paradigm for the geospatial Grid projects being developed.

Figure 1 illustrates the remote sensing data processing with a simple case. Based on the analysis in the last section, our dynamic Grid workflow web portal has been designed
to provide a kind of graphical flowcharts as the raw material to represent the data processing methods, where we use a black-box with many different parameter sets to express the tasks executed and use arcs to represent the relations of the execution order. The parameter sets include control-flow elements such as branching and loops, which provides a dynamic control way for users; and the overall execution elements, which restrict the sequential or parallel execution way of tasks. The transfer of data file and control flow can be accomplished by input ports and output ports. Taken together, workflows, black-box, parameter sets, input ports, output ports and connections, represent the basic function.

The GUI web portal can be represented as a 8-tuples GWP = (K,D,R,P,s,d,F,T), where:

- K is a finite set of states, where each black-box is regarded as a state.
- D is a set of data, which includes the initial data file of remote sensing information, the results of data disposed.
- R is a subset of binary relation K×K. It is a set of arcs, where each element represents the order relation among the executed tasks.
- P is a finite set of parameters, which include the parameters provided by users and control tokens.
- s ∈ K and d ∈ D are the initial state and initial data file.
- F ⊆ K is the set of final states, and
- T is a transition function from (K-F)×(P∪{∅})×D to K×D.

We now formalize the operation of the model. When the Workflow is started, triple(s,∅,d) (K ⊆ F) × (P∪{∅}) is its initial transition state of T; for all q ∈ (K-F) and p ∈ K, if (q,p) ∈ R, the workflow scans its set of data and its set of parameters, getting relevant d1 ∈ D and p1 ∈ P; then the transition is fired, according to the semantic analysis of the parameters, and changing the state q to the state p and producing the result d2 from d1, until it finds a state pi ∈ F; and it halts. The data file di is the result we need.

4. IMPLEMENTATION

We implemented the GUI described in the previous section (see Figure 2). By the GUI, the users can choose an “icon” from the “tree”, drag the “icon” to pane. After clicking the “link” button and choosing two “icons” from the pane, users can get the relation of parallel or serial between steps. Users can modify dynamically the “black-box” and “line” to represent the execution semantic of workflow, accomplishing the utilization of resources and the sharing of the algorithms of remote sensing data processing and analysis. The workflow will be saved and transferred to the remote sensing information service Grid node platform in XML via Internet. According to the semantic, the platform will execute the workflow and response the results to the users.

5. CONCLUSIONS

This paper introduced the necessities for the dynamic Grid workflow and presented a design of dynamic Grid workflow for remote sensing information service based on the 8-tuples. The method of 8-tuples described the transition, control data and control flow. The states are connected through transition. Control data are data used for workflow management purposes. By using the threads, we can control concurrent executions in a workflow. In our future work, we will provide the service of registry for users. Using Internet browser, users can commit their algorithms and services. According to the service of registry, the Grid workflow system can query and find the available resources on the Internet.
6. ACKNOWLEDGEMENT

This work is an output from the research projects funded by NSFC under Grant Nos. 40471091 and 40671142, “973 Project - Active and passive remote sensing of land surface ecological and environmental parameters” (2007CB714402) by MOST, China and “Multi-sources quantitative remote sensing retrieval and fusion” (KZCX2-YW-313) from CAS, China.

7. REFERENCES


