A Processing Framework of Grid Workflow for Remote Sensing Quantitative Retrieval

Jianwen Ai
Guangdong Electronics Industry Institute Corporation Co., Ltd, GDEII
Dongguan, China
neau_ajw@hotmail.com

Yong Xue
Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
Faculty of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK

***Abstract*—Grid workflow is an intuitive and convenient tool for user to use Grid services. It provides a framework for solving the remote sensing distributable computing problem in an integrated unified processing environment of Grid services. This paper firstly introduces the architecture of Grid workflow processing framework applying for remote sensing quantitative retrieval service. And then it gives the relevant method of Grid Workflow modeling and mapping for remote sensing retrieval service. Then the authors give a case of accomplishing the framework.**

Keywords- ***Workflow, Grid, Remote Sensing, Framework***

I. INTRODUCTION

The remote sensing quantitative retrieval need to manage and process large-scale data sets, and execute experiments on distributed resources. Grid technology provides the possibility to solve the above problems. Although there is a good prospect for Grid technology in remote sensing, however, this potential has yet to be developed. And many problems remain to be solved. The distributed and heterogeneous nature of Grid clusters, such as various hardware platforms and software systems, access and interaction interfaces and data and resource management systems, make the Grid environment difficult to use by the layman, and thus require additional management to coordinate the multiple resources.（Jaeger-Frank et al., 2006） Grid workflow for remote sensing quantitative retrieval is the "bridge" between Grid services and Grid enabled application of remote sensing quantitative retrieval. Workflow for Grid computing integrates distributed data, algorithms and components in a Grid computing environment (Yu and Buyya 2005). It avoids low-level details and hence enables users to focus on higher levels of abstraction at application level (Hernández et al 2004).

II. THE ARCHITECHTURE OF THE PROCESSING FRAMEWORK

The workflow description generated in an XML format drives the workflow management process. A more user-friendly method of workflow management is to integrate the generation of the workflow into a graphical user interface. These graphical user interfaces are implemented mainly as grid portals within the used grid infrastructure(Murugan and Kumar 2012).

Many researchers achieved to a variety of management systems for Grid workflows, each dedicated to a particular application domain and based on concepts and specific models (Bendoukha and Benyettou 2012). However, there are a large number of geospatial operations in numerous applications and it is non-trivial to formalize these operations as workflow components due to various syntactic and semantic mismatches. Unlike business data processing that is control flow centric and requires relative insignificant computational resources, geospatial data processing can be data intensive, computation intensive, and, visualization intensive. It is non-trivial to develop a geospatial Grid workflow system. (Zhang, 2012).

Grid workflow for remote sensing quantitative retrieval is a type of grid workflow system, and is built on the basis of resource and service management of grid system. It provides users a tool to compose application of remote sensing based on Grid service. The customized Grid workflow for remote sensing quantitative retrieval need to run on heterogeneous and distributed Grid resources, in accordance with an ordered processing flow to achieve specific goals. Currently, most existing workflow can define domain-specific workflows, but it is difficult to describe the complex relationship of tight coupling remote sensing algorithms. Therefore, Grid workflow for remote sensing quantitative retrieval needs new architecture and scheduling algorithm due to the characteristics of remote sensing data, special processing algorithms, and high transmission cost. This is a new challenge for Grid workflow for remote sensing quantitative retrieval. The authors design the processing framework of Grid workflow for remote sensing quantitative retrieval based on the reference of the Grid architecture. The framework is basically made up of different parts: Grid workflow visualized composition tool for remote sensing quantitative retrieval, workflow submission, workflow parser, Grid workflow for remote sensing retrieval scheduler, Grid enabled for remote sensing data and algorithms based on integration, Grid enabled for resources packaging and cluster(see Figure 1).

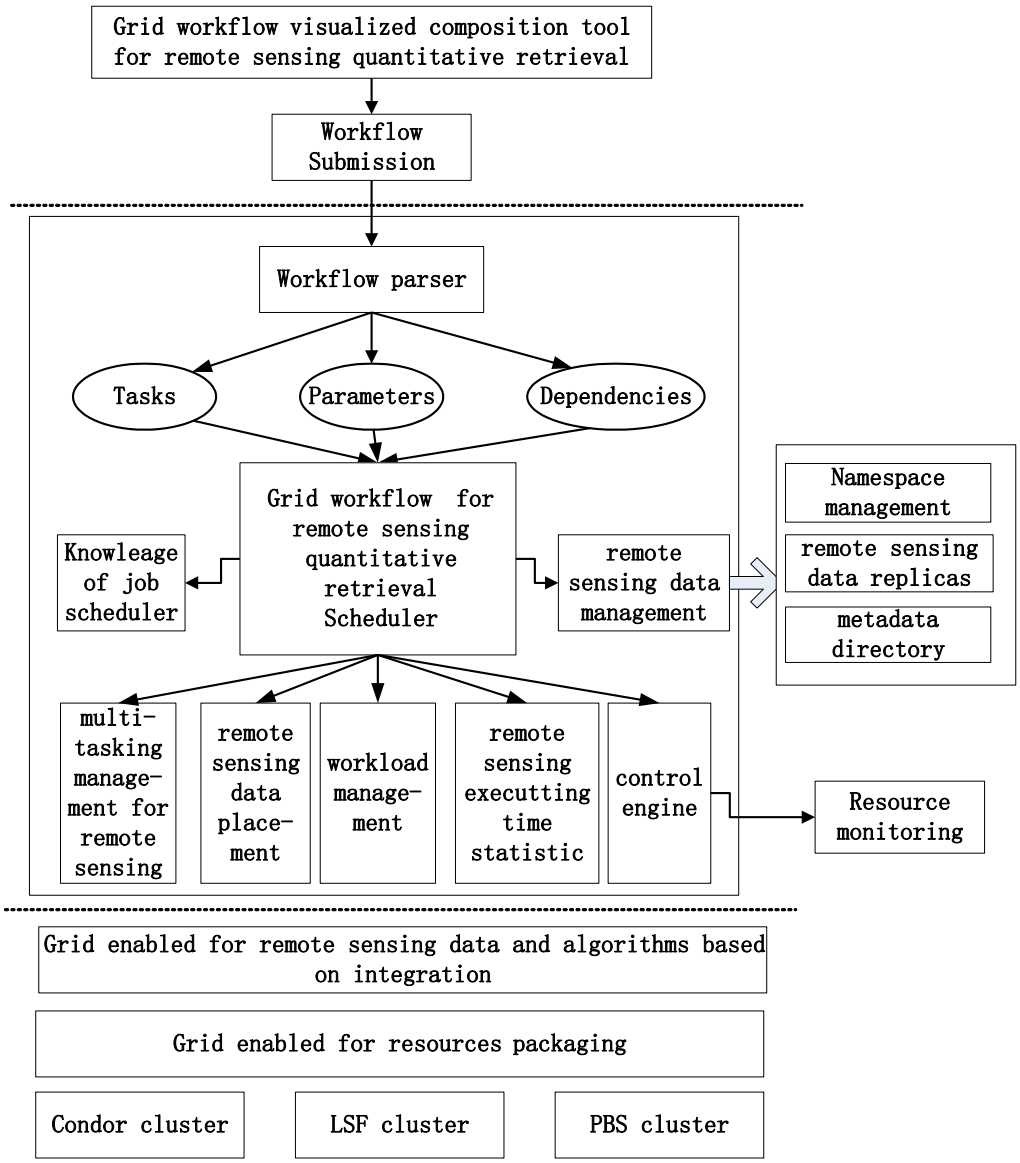

Figure 1. the architechture of the processing framework

As shown in Fig. 1, the framework was designed for parallel job executing among computational nodes. The processing flow is following:

- The user customizes the Grid workflow for remote sensing quantitative retrieval. When composing a case of workflow, the user first composes a workflow process model, and then defines the process model.
- When the workflow is submitted to the task receiver, the workflow parser will analyze it based on the graphics-based analytical workflow. The parser will analyse the customized workflow about its parameters, tasks and relation of dependency among tasks. Then the results will be submitted to the execution engine of Grid workflow for remote sensing quantitative retrieval scheduler.
- The execution engine querys the task scheduling knowledge based on the results of the analysis. Then it will give the scheduling decision-making of the tasks. And it will bundle resources and tasks according to the detected information service.
- When scheduling, the scheduler will allocate the different tasks to Grid infrastructure according to the scheduling decisions, the load and the assessment of resource performance. It will schedule the data transfer engine to achieve massive data transmission. By the controlling engine, it will monitor the Grid resources and the tasks of workflow.
- By using Grid enabled for remote sensing data and algorithms based on integration, Grid enabled for resources packaging and machine resource matching

script. The processing framework will accomplish the mapping from a case of the Grid workflow for remote sensing quantitative retrieval to computing nodes.

III. GRID WORKFLOW MODELLING FOR REMOTE SENSING RETRIEVAL SERVICE

Grid workflow modeling for remote sensing quantitative retrieval service includes that its task definition, structure definition and mapping relation of specific Grid resources for task execution. Task definition includes that all information to execute a task in grid environment, such as function descriptions of task, previous task, support environment requirements, the size of memory, minimum space of hard disk, etc. In general, its workflow structure can be represented as a DAG. The Grid workflow modeling for remote sensing quantitative retrieval illustrates in detail the form workflow elements, including activity, transfer, application and process related data entity definition or declaration. It is the description of the operation state of the workflow system and is the cornerstone of the performance and robustness of the workflow system. The Grid workflow modeling for remote sensing quantitative retrieval is a mathematical method to formally describe the status and changes of the remote sensing quantitative retrieval processing at different moments in the Grid platform. It includes mainly the definition of the routing of remote sensing quantitative retrieval model, the definition of Grid environment of remote sensing quantitative retrieval processing transitions, the mapping of remote sensing quantitative retrieval between model and transitions, the required resources of Grid-enabled transitions, the parameter description of transitions, the description of transitions inspired in Grid environment, the records of relationship among transitions and the relevant data of Grid workflow for remote sensing quantitative retrieval. Its service is a method based on the hybrid Petri nets. Petri nets have a number of major advantages. In the first place, it forces precise definition. Ambiguities, uncertainties, and contradictions are thus prevented, in contrast to many informal diagramming techniques. Secondly, the formalism can be used to argue about processes. It becomes possible to check whether a case is successfully completed after a period of time (Van Der Aalst and Van Hee,2004). Grid workflow modeling for remote sensing quantitative retrieval can be represented as a 7-tuples RSGWF-PN={RSD, RST, S, I,O,C,m_0}, where:

i. RSD ={rsd_1,…,rsd_n},n>0, is a finite set of places;
ii. RST ={rst1,…,rstm},m>0, is a finite set of transitions, with RSD $\cap$ RST = Θ;
iii. S is the status set of current Grid node;
iv. I: RSD$\times$S$\times$RST $\rightarrow$N, N={0,1,…}, is the input function. It defines the set of repeat times or weight for directed arc from RSD and S to RST;
v. O: RST $\times$ RSD $\times$ S $\rightarrow$ N, N={0,1,…}, is the output function. It defines the set of repeat times or weight for directed arc from RST to RSD and S;
vi. M: RSD $\rightarrow$N is a marking where M(RSD) represents the number of tokens in place rsd_i .
vii. C is the color-extended set associating places and transitions. It is used to identify different cases of Grid workflow modeling for remote sensing retrieval. Especially, an initial marking is denoted by m_0 . It represents the initial status of disperse system. Tokens are pictured by dots. It represents the dynamic instance accomplished by places which denotes the partial status.

IV. DESIGN AND MPPING OF GRID WORKFLOW FOR REMOTE SENSING QUANTITATIVE RETRIEVAL

Grid workflow composing system for remote sensing quantitative retrieval is the geospatial problem solving environment portal between remote sensing Grid system and the user. It includes two types of workflow models, namely abstract model and concrete model. The design of Grid workflow composing system for remote sensing quantitative retrieval needs to deal with the following several basic problems:

i. There is visual sequence / routing modeling capability of parallel. It can add, delete and modify routing in the graphical interface. It can separate the element model and the executing example of specific element model. It can express all kinds of relations between various element model by graphical method.
ii. It can describe the routing characteristics of executing tasks, the definition of task to be triggered, the conditions and order of element model for instance needing to be started, the definition of meta model instance to be triggered and the decision-making principle of subsequent element model instance scheduling.
iii. It can achieve the communication with server Web application and obtain the remote sensing quantitative retrieval element model and element model instance tables from the Web server. The scientific users can interactively obtain meta model and its instance description information, input parameters, output parameters, interface description information, control variables and Grid deployment description information.

Grid workflow composing system for remote sensing retrieval provides users with a GUI graphical user interface. It integrates remote sensing algorithms, remote sensing data and large-scale remote sensing data transmission tool together according to the way of workflow. It provides the user an easy-to-use Grid service environment about the remote sensing applications. Using it, the users can express the executing relation between synchronous and asynchronous processing step for remote sensing quantitative retrieval algorithm. It represents the dependencies among the remote sensing quantitative retrieval algorithm steps (see Figure 2).

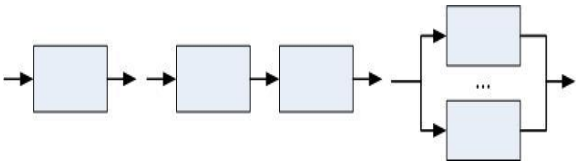

Figure 2. the dependencies among the remote sensing quantitative retrieval algorithm steps

Abstract Grid workflow for remote sensing quantitative retrieval mapping achieves to bind a case of abstract Grid workflow with computing nodes according to the job description file for Grid service. It can be represented as a 4-tuples <RST, Δ, Υ, J>, where:

i. RST is a partial ordering set of remote sensing quantitative retrieval of abstract Grid workflow for remote sensing quantitative retrieval tasks and remote sensing Grid enabled services;

ii. Δ ={(x,y, rst)|x $\in$ {ST $\cup$ DT $\cup$ FT },y $\in$ C_RSD, rst$\in$RST, (x,y,rst) is the valid division of y in the set of rst under the status of x};

iii. J is a Grid job description language or method, including Condor Job ClassAD, SGE script, PBS job management command set, LSF job description script and the European Data grid JDL, GRMS job description, Globus resource specification language RSL and the Global Grid Forum for job submission description language JSDL;

iv. Υ is a mapping between RST and J transfer, which is mainly based on the remote sensing task scheduling knowledge and Grid information discovery service. It determines the mapping RST to J elements.

We have implemented Grid workflow visualized composition tool for remote sensing quantitative retrieval (see Figure 3).

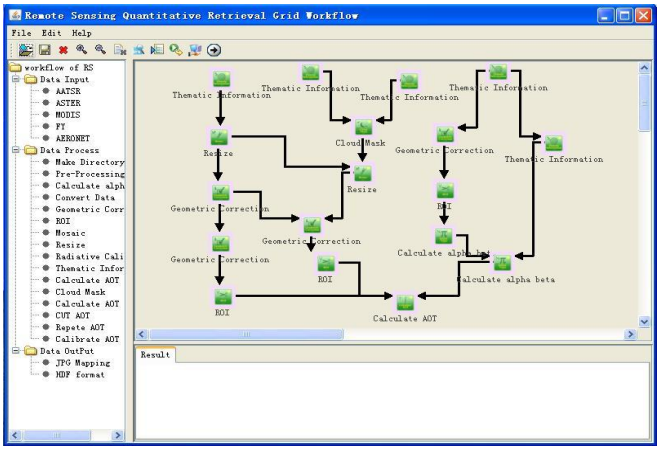

Figure 3. Grid workflow visualized composition tool for remote sensing quantitative retrieval

According to the above method of Grid workflow modelling and mapping for remote sensing quantitative retrieval service, we have implemented it (see Figure 4).

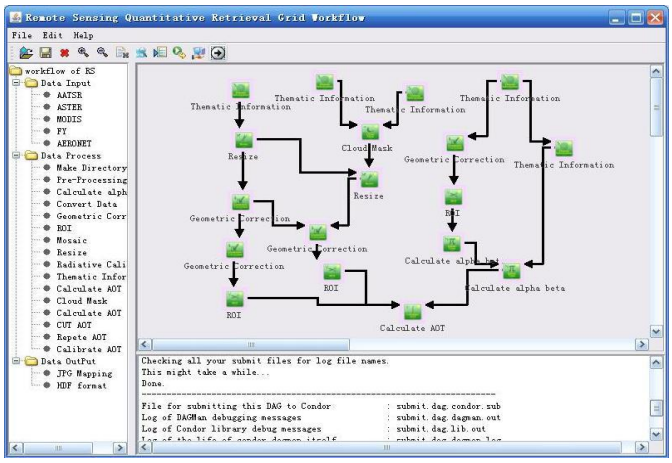

Figure 4. A case of Grid workflow modeling and mapping for remote sensing quantitative retrieval

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 41271371) and Guangdong Province&Chinese Academy of Sciences Comprehensive Strategic Cooperation Project (No. 2011A090100003)as well as Guangdong Province New Strategic Industry Development Special Fund(high-end electronic information of new type) Project (No.2012556028).

REFERENCES

[1] J., Yu, and R., Buyya,2005, "A Taxonomy of Workflow Management Systems for Grid Computing" , Journal of Grid Computing , Vol. 3, Springer Netherlands,pp171-200.

[2] F., Hernández, P., Bangalore, J., Gray and K., Reilly,2004. "A Graphical Modeling Environment for the Generation of Workflows for the Globus Toolkit". In Workshop on Component Models and Systems for Grid Applications,18th Annual ACM International Conference on Supercomputing(ICS 2004),Saint-Malo,France,ACM Press, New York,NY,USA.

[3] E. Jaeger-Frank, C.J. Crosby, A. Memon, V. Nandigam, J R. Arrowsmith, J. Conner, I. Altintas, and C. Baru, Three Tier Architecture for LiDAR Interpolation and Analysis, Lecture Notes in Computer Science on 1st International Workshop on Workflow systems(WSES06),Vol.(3993), pp920-927,2006.

[4] S. Murugan, S. Kumar, A Survey of Workflow Management Tools For Grid Platform[J]. Advances in Information Technology and Management, 2012, 1(1): 1-3.

[5] H. Bendoukha, A. Benyettou, JASMIN: A Visual Framework for Managing Applications in Service-oriented Grid Systems[C]//ICIW 2012, The Seventh International Conference on Internet and Web Applications and Services. 2012: 46-51.

[6] J. Zhang, A practical approach to developing a web-based geospatial workflow composition and execution system[C]//Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications. ACM, 2012: 21.

[7] W. Van Der Aalst, K. M. Van Hee. Workflow management: models, methods, and systems[M]. The MIT press, 2004.