A Model Base Based Grid Workflow Expertise Logical Validation Solution for Quantitative Remote Sensing Retrieval

J. Liu[1, 4], Y. Xue[1, 2, 3], L. L. Liu[3, 4], Z. Q. Chen[3, 4], J. Dong[1,4]

[1] 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, Beijing 100101, China
Telephone: +86-10-64889540
Fax: +86-10-64889570
Email: libraspring@163.com

[2] Faculty of Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
Email: yxue@irsa.ac.cn

[3] Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

[4] University of Chinese Academy of Sciences, Beijing 100039, China

1. Introduction

Advances in sensor technology are revolutionizing the way remotely sensed data are collected, managed and analyzed (Plaza and Chang, 2008), which create higher requirements for computation and resource sharing (Xue et al., 2011), allowing for the larger amount of data, the higher algorithm accuracy and real-time application requirements, redundant general algorithm implementation, etc. Grid computing provides a powerful tool for sharing both remote sensing data and processing middleware (Xue et al., 2005), and is a potential solution for data intensive and complex remote sensing issues (Xue et al., 2010).

The Grid system is used to refer to the software system that provides uniform and location independent access to geographically and organizationally dispersed, heterogeneous resources that are persistent and supported (Johnston, 2001). The remote sensing Grid belongs to the scope of application Grid, while the remote sensing Grid platform involving the workflow technology is a Grid middleware constructed upon the remote sensing Grid infrastructure (Chen and Yang, 2008) and helps to handle complicated remote sensing application including multiple processing algorithms for non-professional users effectively.

In a Grid workflow system, both Grid workflow verification and validation are important. Checking the logical correctness in a workflow consisted of multiple algorithms to obtain meaningful computing results for non-expertise users belongs to Grid workflow validation, which will give instructions to non-expertise users, thus avoiding wasting time and resources.

In this paper we put forward a practical Grid workflow expertise logical validation solution framework for quantitative remote sensing retrieval. This specific solution is implemented on the Grid workflow system - the Remote Sensing Information Service Grid Node (RSIN).

2. The Remote Sensing Information Service Grid Node

The RSIN system based on Condor is built in the TGP laboratory in the Institute of Remote Sensing Applications of Chinese Academy of Sciences, as shown in fig. 1 (Xue et al., 2008).

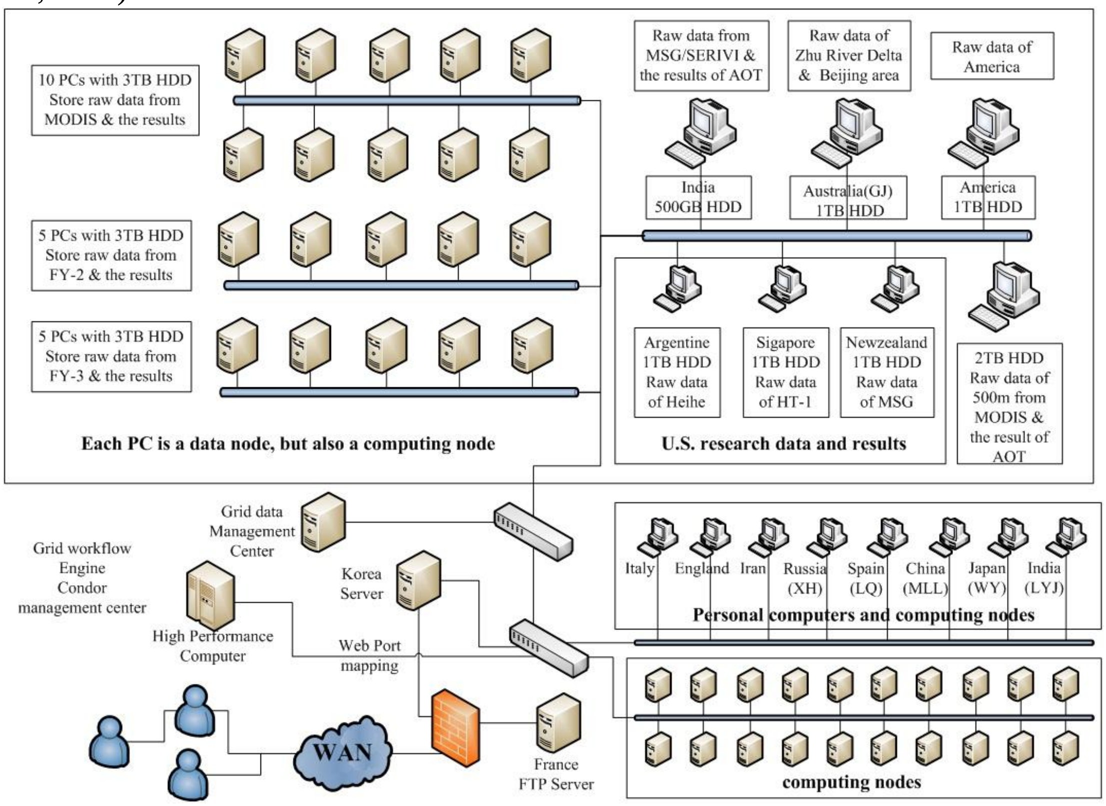

Figure 1. The physical architecture of RSIN system

The four-layer remote sensing Grid architecture includes the Grid base layer, the core Grid component layer, the remote sensing core middleware layer and the application layer. Details about the architecture design and implementation of the RSIN system could be found in the paper by Xue et al. (2011).

3. Grid Workflow Validation

In a Grid workflow system, both Grid workflow verification and validation are important. Verification failure results in the Grid workflow specification and execution containing faults or flaws such as deadlock, temporal violation and resource conflict. Validation failure constitutes a breach of contract between the complex process developer and the client (Kurdel and Sebestyenova, 2008).

To make the RSIN system meet demands of non-expertise users, the Grid workflow validation taxonomy consists of three elements (Chen and Yang, 2008): (a) representation of complex scientific and business processes; (b) representation of expressive power of Grid workflow management systems; and (c) validation approaches. We focus on the middle issue, as currently scientific patterns existing in most Grid workflow management systems are not enough and seldom consider impacts of expertise logical dependencies misuse in specific fields.

4. Expertise Logical Validation Solution

4.1 Dependencies Constraint in Model Base

In a specific application workflow, each model will have its predecessor model except the start ones in every branch of the whole workflow. If a non-expertise user makes a model start before its predecessor in his workflow design, it might lead to meaningless retrieval results or will confront unexpected termination as a result of data unprepared. In a word, there exist dependencies among models in an application workflow, which are similar with the functional dependency in database.

In the relational database theory, given a relation R, a set of attributes X in R is said to functionally determine another set of attributes Y (written $X \rightarrow Y$), if, and only if each X value is associated with one Y value (Halpin and Morgan, 2008). Similarly, we define the dependencies constraint in model base as that if, and only if model X has executed, model Y could generate useful results, then the application workflow is said to satisfy the model dependencies $X \rightarrow Y$. This sort of extension of functional dependency has been applied to other fields such as XML design (Vincent et al., 2004; Chen et al., 2010).

We also extend the axiom of transitivity among Armstrong's axioms (Halpin and Morgan, 2008) and apply it to the RSIN system: If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$, in which X, Y and Z are remote sensing models. Therefore, we could simply assign an additional attribute 'straightforward predecessor' for every model, whose corresponding value is determined by expertise-users, and will join and search in multiple model information lists when necessary to obtain complete dependencies to validate logic by non-expertise users.

4.2 Framework of RSIN System involving Validation Solution

In the framework of RSIN involving validation, we modified the application layer of the original architechture of RSIN by dividing the Graphic User Intrerface (GUI) tool according to two roles: expertise and non-expertise, and making use of the model base in the grid base layer.

The RSIN system authorises expertise-users to log model informations and determine dependencies constraint of multiple algorithms in model base. When non-expertise users design a application workflow with the GUI tool, it will give instructions in the light of informations stored in model base. If users compulsorily execute the workflow despite of existing conflicts between the design with the dependencies constraint in model base, the system will not terminate the execution but train informations in model base to meet specific requirements of the coresponding users. The expected process of the RSIN system is shown in fig. 2. We initially construct a prototype of the modified GUI tool, as shown in fig. 3 and fig. 4. The fig.3 represents interfaces for the login and expertise users, and the interface in fig. 4 will exist after the non-expertise users login.

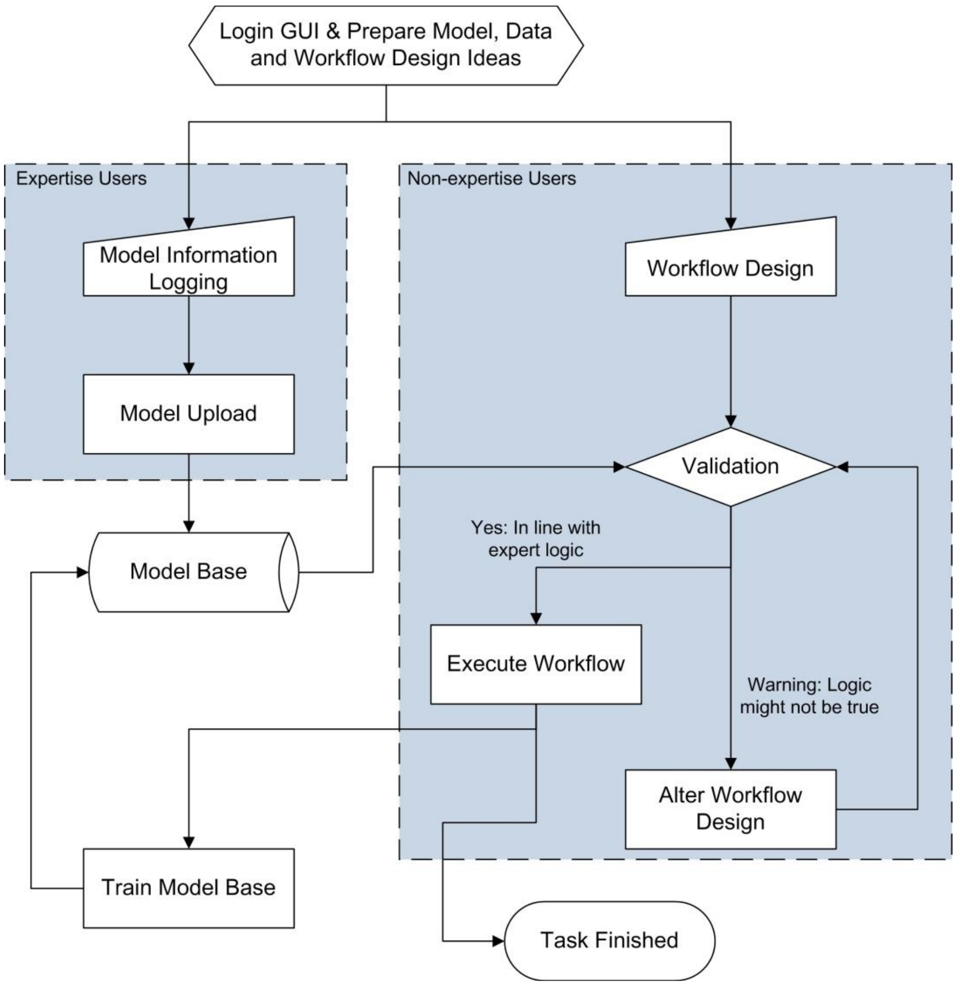

Figure 2. The flow scheme of RSIN involved with validation.

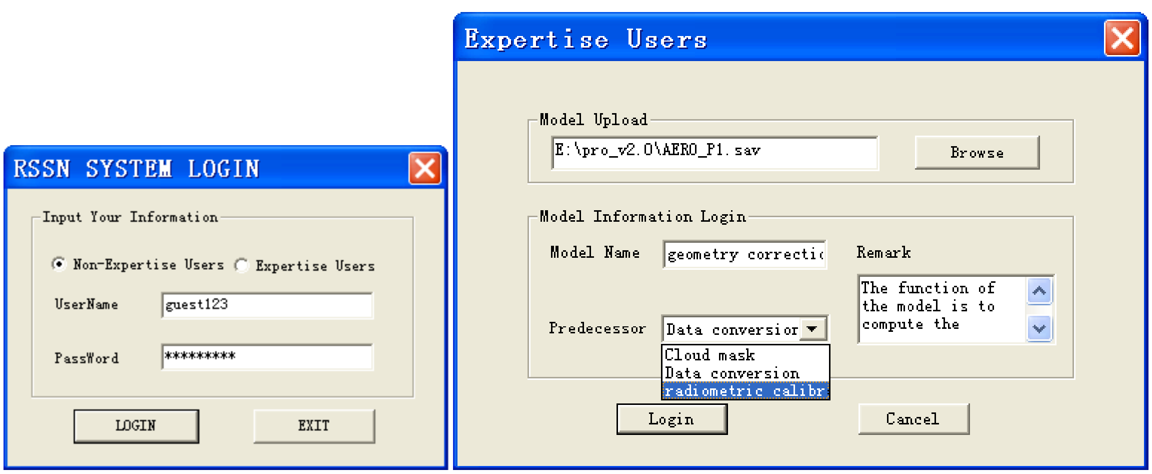

Figure 3. The interfaces for the login and expertise users.

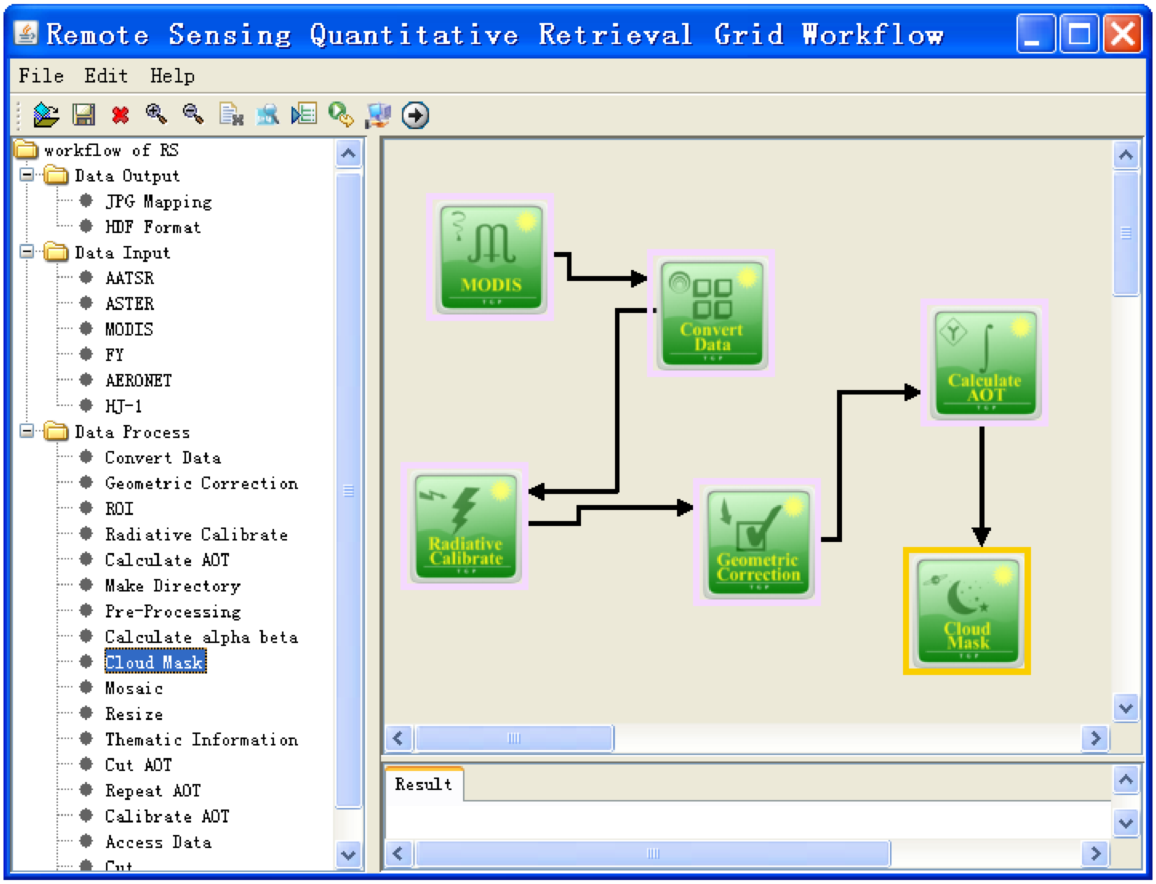

Figure 4. The interface for non-expertise users.

5. Conclusions

The RSIN involving the expertise logical validation will not only accelerate the quantitative remote sensing retrieval via utilizing Grid resources but also facilitate the application construct through the GUI tool as well as offer constraints to avoid wasting time and resource. The validation way proposed in this paper is a simple but practical solution. After the modified GUI tool is completely improved and tested, we would apply it to our typical AOD retrieval and more potential remote sensing scientific applications.

6. Acknowledgements

This work was partly supported by the Ministry of Science and Technology (MOST) of China under Grant No. 2010CB950803, 2010CB950802 and by NSFC, China under Grants 41101323.

7. References

Chen HT, Liao HSand Gao ZQ, 2010, Functional dependencies for XML. *Lecture Notes in Computer Science*, 6185: 110-115.

Chen J and Yang Y, 2008, A taxonomy of Grid workflow validation and validation. *Concurrency and Computation: Practice and Experience*, 20 (4): 347–360.

Halpin T and Morgan T, 2008, *Information modeling and relational databases*. Morgan Kaufmann, San Francisco, USA.

Johnston WE, 2001, Using computing and data grids for large-scale science and engineering. *International Journal of High Performance Computing Applications*, 15(3): 223-242.

Kurdel P and Sebestyenova J, 2008, Grid workflow specification and verification. *WSEAS Transactions on Computers*, 8(7): 1199-1208.

Plaza AJ and Chang CI, 2008, *High performance computing in remote sensing*. Chapman & Hall/CRC, Boca Raton, FL, USA.

Vincent IW, Liu JX, Liu CF, 2004, Strong functional dependencies and their application to normal forms in XML. *ACM Transactions on Database Systems*, 29(3): 445–462.

Xue Y, Ai JW, Wan W, Guo HD, Li YJ, Wang Y, Guang J, Mei LL, Xu H, 2011, Grid-enabled high-performance quantitative aerosol retrieval from remotely sensed data. *Computers & Geosciences*, 37: 202-206.

Xue Y, Ai JW, Wan W, Li YJ, Wang Y, Guang J, Mei LL, Xu H, Li Q, Bai LY, 2010, Workload and task management of Grid- enabled quantitative aerosol retrieval from remotely-sensed data. *Future Generation Computer Systems*, 20: 590-598.

Xue Y, Chen ZQ, Xu H, Ai JW, Jiang SZ, Li YJ, Wang Y, Guang J, Mei LL, Jiao XJ, He XW, Hou TT, 2011, A high throughput geocomputing system for remote sensing quantitative retrieval and a case study. *International Journal of Applied Earth Observation and Geoinformation*, 13: 902-911.

Xue Y, Palmer-Brown D, Guo HD, 2011, The use of high performance and high throughput computing for the fertilization of digital earth and global change study. *International Journal of Digital Earth*, 4(3): 185-210.

Xue Y, Wan W, Li YJ, Guang J, Bai LY, Wang Y, Ai JW, 2008, A data intensive scientific computing framework for quantitative retrieval of geophysical parameters using satellite data. *IEEE Computer*, 41(4): 33-40.

Xue Y, Wang JQ, Wang YG, Wu CL and Luo Y, 2005, Preliminary study of grid computing for remotely sensed information. *International Journal of Remote Sensing*, 26(16): 3613-3630.