Distributed software framework and continuous integration in hydroinformatics systems

Jianzhong Zhou¹, Wei Zhang¹, Mengfei Xie¹, Chengwei Lu¹, Xiao Chen²

¹School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan China
²School of Management, Huazhong University of Science and Technology, Wuhan China

Abstract: When encountering multiple and complicated models, multisource structured and unstructured data, complex requirements analysis, the platform design and integration of hydroinformatics systems become a challenge. To properly solve these problems, we describe a distributed software framework and it’s continuous integration process in hydroinformatics systems. This distributed framework mainly consists of server cluster for models, distributed database, GIS (Geographic Information System) servers, master node and clients. Based on it, a GIS - based decision support system for joint regulating of water quantity and water quality of group lakes in Wuhan China is established.

1 Instructions
In recent years, to meet the requirements of water resources informatization and management, a variety of hydroinformatics systems covering different subject are developed [4, 5] and more recently [7] respectively established their systems for flood management and related evacuation planning. Applications released by Alemu et al.[1], Booth et al.[2], Zhang et al.[8], focusing on water quality monitoring or optimizing reservoir operations, are also highly successful in their field. However, when an information system (IS) or decision support system (DSS) is involved with complex models and data from different fields (hydrology, hydrodynamics, risk assessment), the process of system development become a challenge. Even the initial deployment and subsequent dynamic reconfiguration of such a software system are time-consuming and tedious.

As the size and complexity of software systems increases, procedures (manual or automated) that assume a particular software architecture and environment are becoming untenable. Therefore distributed software framework and related applications were developed to replace traditional centralization methods not only in development but also in deployment. Exactly, comparing with common software system, the advantages of distributed system line in disreputability, communicativeness, robustness and loose-coupling, which can technically support the better performance of inter-discipliner models and convenient storage of heterogeneous data in hydroinformatics. Thus, applying distributed software framework to application design and development is a feasible solution when establishing complex hydroinformatics system.

With the rapid expansion of agile development, continuous integration (CI) process during software development has received wide attention. CI originated from the theory of Extreme Programming (XP) in agile development [2]. It encourages developers to integrate their application components frequently for quality improvement and risk reduction. In accordance with the definition of Open
Complex Giant System (OCGS) proposed by XS Qian, a hydroinformatics system, which covering multi-subjects and multi-businesses, is also a OCGS [6]. It demands Iterative development, continuous deployment and continuous test, therefore it’s continuous integration is also emphatically discussed in following paper.

2 Distributed Software Framework

Distributed Software Systems is defined as a type of software system that support distributing processing, distributed software system can be operated based on components located on networked computers, communicate and coordinate their actions by passing messages [3]. This system contains three basic elements: distributed operating system, distributed programming and compiling system, distributed documents and database. Hydroinformatics is well known as a cross-disciplinary research topic which is quite relevant to the application of GIS. The traditional software framework that connotes three layers (Data access Layer, Business and logic layer, Presentation layer) cannot meet actual software development need. Therefore, we introduced the distributed software system(Figure 1), a more complex software developing framework that also consists of three layers:

1) Basic layer: The main components of the basic layer are: distributed database, cluster of server for models, GIS server. a) distributed database, in addition to the traditional database that maintains the connected database, the database of distributed systems integrates two new database to the systems, the geodatabase that support geodata processing and the file database that support unstructured and half-structured data processing. b) cluster of servers for models, in order to achieve efficiently distributed calculation and meet system operating need derive from different platforms or environments, the distributed system also integrate the complex hydroinformatics models to the server clusters for models, such as: hydrodynamic model, hydrologic model, reservoir Group operation model. c) GIS servers, the operation of GIS is also allowed in the basic layer, because GIS is highly relevant for studying hydroinformatics models, the usage of GIS are mainly related to pre- or after-processing of data in the models, visualizing the results, publishing the map service.

![Distributed Software Framework Diagram](image)

Figure 1 Distributed software framework for hydroinformatics systems

2) Service Layer: The main object of service layer is an autonomous working server called Master Node. It is responsible for remote access services management, which mainly contain model computing service, data access service, GIS-based data conversion service and so on. All these remote services are registered in a service library stored in the local database. Additionally, some relatively
sample models (evaluation models, statistical models), which not require specific running environment or platform, can be integrated in this master node. Furthermore, it also playing the role of controller for data transmission and management among three units in basic layer. The internet between basic layer and service layer is supposed to be private network while the service request interface should be exposed for public accessing.

3) Presentation Layer: Presentation layer is composed of various clients (universal client or recommended client). Different clients and user groups may have different permissions for service access. For enhancing the universality of whole system, clients can be respectively established on different operating systems (Windows, Linux, MacOS), running environments (.Net, JRE) or devices (mobile phone, PC).

3 Continuous Integration
In software engineering, continuous integration is the automatic build practice where team members could integrate their work frequently, which usually means that each person will integrates at least one time within one developing cycle (daily or weekly). In purpose of detecting integration errors as soon as possible, each integration is supposed to accomplish an automatic build and function testing. Based on continuous build, continuous testing and continuous feedback, adoption of CI can significantly reduced integration problems and it allows a team to develop complex software systems more rapidly.

In high cohesive software systems, CI always mean an integration tool or environment (Jenkins, Hudson, etc) for daily build. However, in the development of large-scale distributed systems, CI does not mean an specific tools any more, but an iterative optimization development process. CI process, which matched with above-mentioned hydroinformatics oriented distributed software framework, is shown in Figure 2. It is an iterative loop which mainly consists of five stages: continuous design, continuous development, continuous deployment, continuous testing and continuous feedback. Continuous design refers to the rising practice of new decision-making ideas and practical requirements from practitioner or decision-maker. Related model (mathematical model or GIS model) modification or construction will be registered and reported to master node by researchers from different disciplines. At the same time, the new model will be deployed to appropriate server with specific service interface and clients are updated to adapt these changes. Therefore, feedback related new functional components, services, or clients are released to meet their needs. Then it is time to proceed the new testing and assessment of whole software system and generate a new feedback report of it.

The continuous testing will go through three stages, unit testing, integration testing and systems testing. Unit testing is referred as module testing as well, among the three-layer structure of the distributed system, the content of unit testing is to test the accuracy and stability of following model related procedures: mathematic calculation, service interface, and data transmission. In practical unit testing, even though it seems like that different unit testing are independent to each other, when encountering complex task, certain testing would need to coordinate different units testing to verify the accuracy and stability. For example, when running the flood simulation for certain river or plane, this function might involve several models, such as 1D and 2D coupled hydrodynamic model, time series based data model, and visualization related GIS conversion model. Therefore, the unit testing would need cooperation between model controller, database administrator, and GIS technicist. Integration testing works based on basic layers and service layers, after the accomplishment of the unit testing, the system will register all the model services in one task category to the master node, which will achieve the management of remote service access. The content of integration testing is to make sure that under remote access working mode, multiple model services would maintain functionally coupling work. System testing is a full link test between user interface and bottom model services and database. Other than technically fixing the errors and bugs in client or system, for better functioning, the feedback reports based on user experience would also be generated after the system testing.
4 East Lake Groups DSS for Joint Regulating of Water Diversion

The demonstrating project applying the DSS is the Water quality and quantity regulation project of East lake groups. East lake groups is located in Wuhan city, the south side of the Yangtze river, it is well known as the largest urban lake groups composing of six lakes linking by 14 tunnels (East lake, Sha lake, Yangchun lake, Yanxi lake, Yandong lake, North Lake), the lake groups come across three districts of Wuhan city, Hongshan district, Qingshan district and Wuchang district. After combining the lakes and involved districts, we got the surface area of 390.6 km² as our final studying area. With rapidly increase in population and speedily development of economy, this lake groups receive much more urban discharge than before, which leads to serious eutrophication and water quality deterioration. Therefore, we propose that the core purpose of this project is to achieve water diversion from the Yangtze River to lakes, by utilizing optimized joint water resource regulating method, we hope to achieve better control of water pollution and recovery of urban water environment of this lake groups. Based on above statements, a GIS-based East lake groups DSS for joint regulating of water quantity and water quality is established.

Development of our DSS is based on WPF(Windows Presentation Foundation) for application user and Silverlight for web user. The GIS module is ArcEngine (ArcGIS Engine) in server side and ArcGIS runtime for .Net in client side. The prototype of DSS is shown in Figure.3. Based on above-mentioned three layers distributed framework, several hydrological and hydrodynamic models are deployed in server cluster to support the asynchronous and parallel mathematical computing of whole system, while models for multi-scheme assessment are integrated in master node. In the aspect of distributed databases, we adopt the Sqlserver 2008 R2 for traditional structured data storage and ArcSDE for Geodata management. Visualization related tasks such as raster layer rendering and vector layer drawing in GIS are integrated as geodata conversion models in two GIS servers.

Our DSS can be divided into seven modules(Real-time monitoring, Digital lakes, Lake hydrology, Flow fields, Pollutants migration, Joint regulating, Assessment). The function of real-time monitoring module is to measure and record rainfalls, water temperature, water flow and state of pumping station. The distributed raster model of Xinan river which is based on history data of rainfall is integrated to the lake hydrology module. Through combining the hydrological module and the history rainfall data, we can run the simulation of flow concentration and non-point source pollution for the study area. Flow fields and pollutants migration are two hydrodynamic related modules, by evaluating the interactive function of wind field and various water regulating plans on the lake, we can accomplish flow field analysis, backwater analysis and pollutants migration simulation by coupling applica-
tion of hydrodynamic model and water quality model. Joint regulating module is the core element of DSS, based on jointly analyzing multiple water diversion route, amount of water diversion, regulating dispatch and final water quality improvement, this module will generate thousands of final regulating plans and related regulating results. Through constraining the regulating goals, decision maker can acquire customized regulating plan satisfying their certain requirements. The last module of DSS is the assessment module, this module will rate the improvement utility and economic cost of a regulating plan, at the same time it allows comparison of various index of multiple regulating plans.

This project, which covers an intelligent decision-making process from the very first rainfall monitoring, non-point source pollution analysis, pollutants migration simulation, scheme generation of water diversion to final multi-scheme comparison, can introduce users who do not have sufficient knowledge of hyinformatics to understand the principles of water diversion and it’s utility.

Figure 3 Prototype of East lake groups DSS

5 Conclusion

This work describes the application of distributed software framework in complex hydroinformatics systems, which can help developers and researchers from different sectors to manage their models, share their data and accomplish coordinating activities for coupling tasks. Distributed framework has strong robustness, flexibility, and interoperability, which also aids in maintainability and extendibility of large-scale software systems. Other than this, the CI process for building distributed system is also emphatically discussed in this paper. In accordance of the iterative developing loop in CI, continuous improvement and continuous feedback are two core properties when conducting distributed development. Depending on it, developers, managers, practitioners and decision-makers can efficiently deliver their message and rapidly correct the errors in system. Based on these methodologies, a GIS-based DSS for joint regulating of water diversion of East lake groups in Wuhan is established, which covering the subjects of hydrology, hydrodynamics, Optimal scheduling and assessments. With it, decision-
maker can acquire real-time information and reliable simulation results of the water diversion project in East lake groups.

Acknowledgment
This work is supported by the National Natural Science Foundation Project of China (NSFC) (Nos. 51239004 and 51579107).

References
[1] Alemu, E. T., Palmer, R. N., Polebitski, A. & Meaker, B. 2010. Decision support system for optimizing reservoir operations using ensemble streamflow predictions. Journal of Water Resources Planning and Management, 137, 72-82.
[2] Beck, K. 2000. Extreme programming explained: embrace change, addison-wesley professional.
[3] Coulouris, G. F., Dollimore, J. & Kindberg, T. 2005. Distributed systems: concepts and design, pearson education.
[4] Levy, J. K. 2005. Multiple criteria decision making and decision support systems for flood risk management. Stochastic Environmental Research and Risk Assessment, 19, 438-447.
[5] Liu, Y., Zhou, J., Song, L., Zou, Q., Guo, J. & Wang, Y. 2014. Efficient GIS-based model-driven method for flood risk management and its application in central China. Natural Hazards and Earth System Science, 14, 331-346.
[6] Xuesen, Q., Jingyuan, Y. & Ruwei, D. 1993. A new discipline of science—The study of open complex giant system and its methodology. Journal of Systems Engineering and Electronics, 4, 2-12.
[7] Zhang, W., Zhou, J., Liu, Y., Chen, X. & Wang, C. 2016. Emergency evacuation planning against dike-break flood: a GIS-based DSS for flood detention basin of Jingjiang in central China. Natural Hazards, 81, 1283-1301.
[8] Zhang, X., Huang, G. H., Nie, X. & Lin, Q. 2011. Model-based decision support system for water quality management under hybrid uncertainty. Expert Systems with Applications, 38, 2809-2816.