International Conference on Hydroinformatics

2014

The Hydro-Modeling Platform (HydroMP) - Enabling Cloud-Based Environmental Modeling Using Software-As-A-Service (SaaS) Cloud Computing

Ronghua Liu

Jiahua Wei

Haiyan Liu

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

THE HYDRO-MODELING PLATFORM (HYDROMP) - ENABLING CLOUD-BASED ENVIRONMENTAL MODELING USING SOFTWARE-AS-A-SERVICE (SAAS) CLOUD COMPUTING

RONGHUA LIU (1), JIAHUA WEI (1), HAIYAN LIU (2)
(1): State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China
(2): Department of Computer Engineering, Beijing Information Technology College, Beijing, 100016, China

Hydro-model has become important tool for water resources management, with higher demand in simulation precision and speed of decision support, models designed for sectoral application becoming outmoded and original mode that massive schemes are run sequentially cannot meet the real-time requirements, especially with the computation increase by finer discretization granularity and broader research range. Water management organizations are increasingly looking for new generation tools that allow integration across domains, and can provide extensible computing resources to assist their decision making processes. In response to this need, a hydro-modeling platform(HydroMP) based on cloud computing is designed and implemented, which can deployed in distributed HPC Cluster and center HPC Cluster use a resources balancer to manage load balancing. This platform integrates multi models and computing resources (i.e. blade computer) dynamically to assure models integrated in platform get extensible computing capacity. A server, hosting HydroMP Web Service and interfaces, is connected to the HPC Cluster and Internet constituting the gateway for registered users. Any terminal (i.e. decision making system) can reference library and Web service of HydroMP in their systems. Massive modeling schemes can be submitted by different users simultaneously, and terminal can get simulation results from HydroMP real-time. Some key approaches and techniques are utilized including: i) a standard model component wrapper communicating with platform by named pipe have developed. OpenMI-compliant model-components can be integrated to this wrapper; ii) API and Event-Handler interface provided by HPC Server, task scheduler and calculation management table is employed to dispatch computing resource, while controlling multiple concurrent scheme submitting; iii) Interface array(i.e. SchemesSubmit, StatusInquiry, GetResult) in the Web Service is supplied to make terminal communicate with platform; iv) Oracle database is used to manage massive model data, results and model-components. This paper describes the details of design and implementation, and gives a case presentation platform application.

Keywords: hydro-modeling platform; model integration; cloud computing; Software-as-a-Service; HPC Cluster

Introduction

Modern water management decisions are increasingly dependent on efficient numerical simulations with multiple models and multiple scenarios. Decision support requirement changes lead to increasing resource consumption and prolonging simulation calculation time, and there is a great contradiction between the simulation precision and timeliness of water management decisions; On the other hand, multiple scenarios comparison and selection of multiple models for result checking are in favor of the reliability of decision, but also faces the contradiction of timeliness and dynamically model integration. The development of cloud computing provides new ideas and pattern for a hydrodynamic simulation service.

Regarding the advantages of cloud computing, in the last two years many scholars and relevant agencies have conducted studies on simulation services and integration platforms based on cloud computing. Sun (2013) used Environmental Decision Support System (EDSS) tools in the cloud services provided by Google Drive to solve high cost problems, to ease difficulty in information sharing and other problems in joint decision-making, and also promoted collaborative decision-making among various inter-basin agencies and stakeholders. Burger et al (2012) applied the concept of cloud computing to integrate the computing power of supercomputer and hydrological models, and used GUI and other interfaces to provide the users with real-time simulation services of the ParFlow hydrological model. Brooking and Hunter (2013) developed a Web-based repository to provide high-speed, interactive access to online simulations of hydrological models. This system stores and renders the schemes for which the computation has been completed, and uses the interface of the Web browser to allow all registered users to inquire and compare these schemes and the simulation results.

Based on the concept of cloud computing, the pattern of hydrodynamic simulation cloud service for the demand of multiple models dynamic integration and calculation resource of high consumption is proposed, and the HydroMP(Hydrologic/Hydraulic Modeling Platform) based on cloud computing is designed and developed.

Framework of HydroMP

Adopting distributed deployment structure, the platform realized the rapid expansion of cluster computing resources capacity; two common methods of model integration are used to realize the model of ‘plug and play’; some assemblies based on oriented object program is designed for data efficient transmission and simulation data structures extensible; Through API provided by the Windows HPC Pack, an efficient integration of the HPC Cluster, hydro-models and platform was realized. The Web Service interface was used to provide online simulation service, to take the efficient hydrodynamic models and computing resources to the terminal. The contradiction of high cost and timeliness of simulation in traditional mode is solved. In traditional model the terminal can use only the local computing model and computing resources.

There is big difference in framework, model invoking method, resources use mode and communication between model and software system in HydroMP architecture, comparing with typical modelling software (i.e., software of desktop version). Modelling schemes and simulation result are stored in a database and multi-models can be dynamically integrated in HydroMP. Models integrated with computing resources serve for hydraulic simulation via web service interface. The HydroMP comprises two parts of cloud and terminal in software framework, the cloud part includes data, model and computing resources, communicate with terminal via web. The terminal and cloud is located in different LAN, and terminal maybe PC Desktop client, Mobile client, Pc Web Browser, Pad Client, and application system across OS and network (i.e. decision support system).

The relational database is used for storage and management of user data, scheme data, computing result data and model information in data layer (Horsburgh, et al., 2008), where tables are linked using the foreign key. Each scheme can be converted into a HydroML file which is a file formatted in XML. Multi-models registered and managed by system in model layer combined into a model base. In model layer, the directory and name of model execute file and data exchange file are stored in the model information table. The management & service layer is responsible for base data management, model management, communication between model and system, scheme data management, result query and system management, the scheme computing function and result query function is encapsulate into web service interfaces to be invoked by any terminal. The framework of system is show in Figure 1.

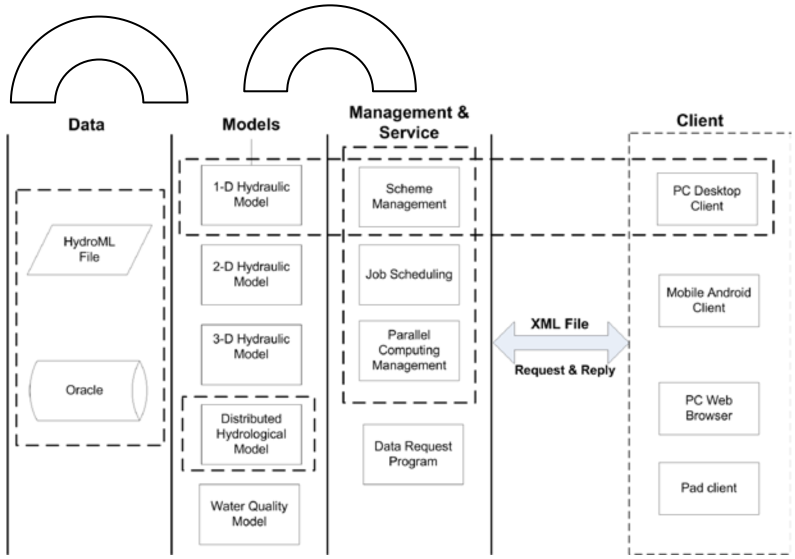

Figure 1. Framework of platform

Distributed deployed architecture

HydroMP is designed into a distributed deployed framework for extension of computing resources and load capacity. Each HydroMP Cluster consists of a HydroMP Center and many HydroMP Servers. Each HydroMP Server can be served for internal users as well as one computing node in the HydroMP Cluster through allocate some idle computing power for public service, in this way the HydroMP is a hybrid cloud service mode. HydroMP Center achieved load balance in the HydroMP Servers based on the PSO and the simulation cost correction stage by stage.

Based on elastic partition, simulation process controller and the state management cache table, HydroMP Server achieved resource scheduling and concurrency scenarios management. The simulation process controller is designed to control the workflow of each simulation scheme, which receives the string bytes from platform, and then call the API provided by Windows HPC Server to create a job and task to submit the computation with the hydro-model and simulation data in HPC cluster. In the computation process, the controller tracks the computation status of HPC job using the Event-Handler interface which is set after the submission of HPC job.

Criteria for resource scheduling include urgency degree and user level. To assure the computation of urgent scheme without increasing the computing resources cost distinctly, method of task pause and restart is proposed in this paper based on feature of the time sequence model, that the next time step initialization condition is the result of back time step. Figure 2 shows the program flow of computation scheduler.

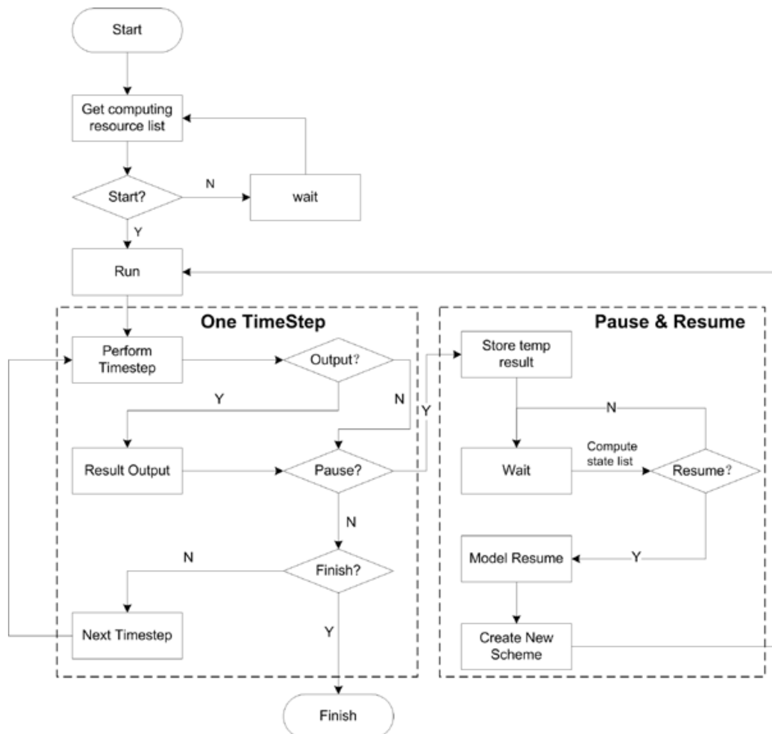

Figure 2. Program Flow of HydroMP computation scheduler

Model integration

Two model integration approaches, including EXE integration and interactive integration, were proposed. The EXE approach was used to realize rapid legacy model integration. The interactive integration method based on the OpenMI concept using named pipe communication and model wrapper ensures real-time communication between the model and the platform during simulation process.

Loosely coupled communications between the platform and the models for integration is proposed, which name EXE integration mode in HydroMP. In this way the data I/O converter established between the platform and the models conducts the data format conversion, and the communication between the platform and the models is realized via size of output file for interaction. Another model integration mode is interactive integration mode (IIM), in which mode, a standardized model wrapper program is developed for communication between model process and computing scheduler of platform via named pipes, and five standardized communication interfaces is used to implement communication. In simulation process, the computing scheduler sends all requests, and the standardized wrapper program responds to the requests and returns corresponding data, thereby realizing the control of the simulation workflow by the platform. Inside the wrapper program, model initialization, preprocessing, simulation startup, single-step boundary updates, single-step execution, calculation completion and other interfaces are used to call methods in the model components, as illustrated in Figure 3.

The standard model wrapper program refers two DLL files including the model component and data exchange file in the wrapper model process. Model component is generated from source model code to compliant to the program and workflow of standard model wrapper, including the model input, mode output and boundary condition of singe time step. The data exchange interface function to exchange model interactive data in each flow into string byte stream. The setting of boundary condition in each time step makes the model wrapper compliant the interaction between multi-models and each model component compliant the OpenMI specification can be integrated into platform.

Through adding the model component and data exchange DLL file as program reference the wrapper program can be rebuilt to a new model EXE file. The new exe file can be registered into platform as a registered model, and invoked by computation scheduler dynamically in simulation process without recompiling of platform. Using the proposed two model integration method, CE-QUAL-RIV1 model, JPWSPC-SC model and JPWSPC-PC model are integrated into platform.

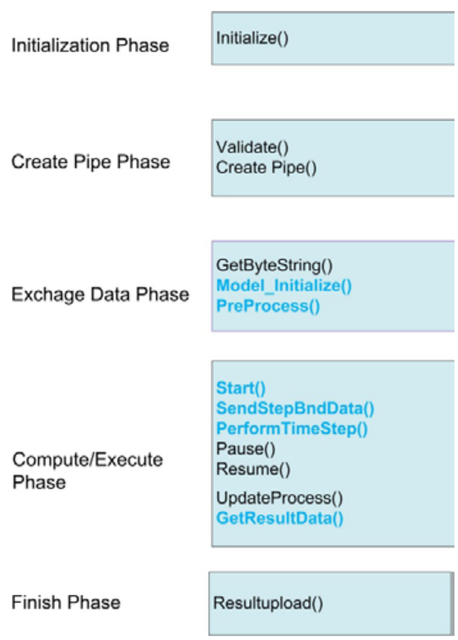

Figure 3. Execution process in the SPIIM integration mode

Case Study

The South-to-North Water Diversion Project (SNWD) is utilized to test the platform, which is a large-scale water diversion project in China to remit the water resources shortage in North China. The flow in channel is fluctuant according the change of head works and bleeders. The discharge change of any bleeder will lead to the fluctuation of the respective channel segments so far as to the whole channel. The start and stop control of gate in the fluctuant process is a complex issue for stable run of SNWD.

Some extreme operation condition is modeling and simulated in this platform. The discharge of head works change in a linear mode and the drainage water volume of each bleeder changes in respective mode. To response to the change of head works and bleeder, the gatage need to be adjusted dynamically to keep the water level change of each gate in a stable range. The five operation works designed in some extreme operation condition is show in table 1, where the discharge change ratio means the discharge change of head works relative to the design flow between start time and end time, the time cost for change mean the allowable time cost to restore stable state of channel.

To get the optimal adjust process of each gate for lesser change of water level of all gates, multi-scenarios required to created, simulated and compared in the platform.

Table1. Discharge change curve in various scenarios

Scenario No	Discharge change ratio (%)		Time cost for change(min)	
	Start time	End Time	Gatage	Discharge of bleeder
1	70	50	10	10
2	70	10	30	30
3	70	80	10	10
4	10	70	30	30
5	70	10	30	30

Multi-scenarios were submitted and computed concurrently in HydroMP and the simulation result could be got in the mode of result increment stage by stage. The channel control system adjusts the scenario variables to optimize the adjustment scheme synchronously using the IIM integrated models (e.g. JPWSPC-SC, JPWSPC-PC). The figure 4 shows the water level change process of some gate in a relative optimal adjust scheme for scenario 1.

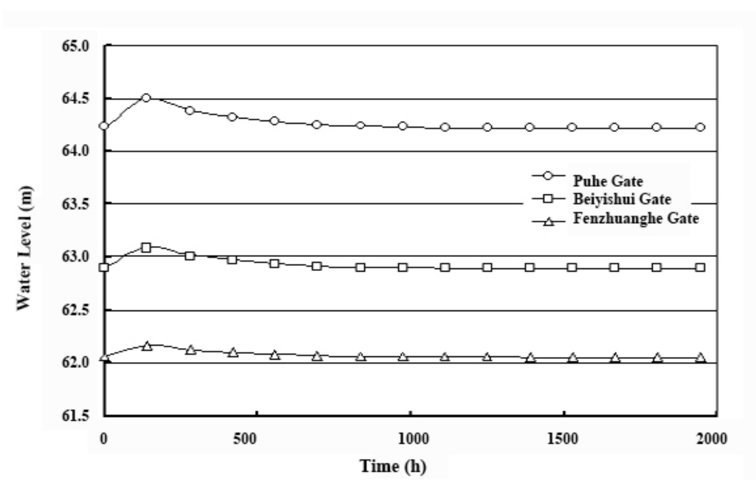

Figure 4. Water level change of some gates

The computation cost in an optimization iteration of each scenario is shown in table 2. The average computation cost using the parallel model (JPWSPC-PC) with 14 processes is 57s, relative to 407s using the JPWSPC-SC in sequential mode. The average speedup ratio is 7.01.

Moreover, the speedup ratio in the concurrent computation mode using the parallel model in the HydroMP is 23.5, comparing to the sequential simulation in sequence on single server.

Table2. Time cost and speedup ratio in cloud computing platform

Scenario No	Used model	Model type	time cost (s)	Speedup ratio
Scenario1	JPWSPC-SC	Serial model	415	-
Scenario2	JPWSPC-SC	Serial model	400	-
Scenario3	JPWSPC-MPI	Parallel model	56	6.9
Scenario4	JPWSPC-MPI	Parallel model	54	7.87
Scenario5	JPWSPC-MPI	Parallel model	61	6.26

Conclusion

Using the relational database, common model integration methods, distributed deployed architecture, as well as other technologies the hydraulic simulation platform based on cloud computing is constructed and three models have been integrated into the platform. Through platform test of the case of SNWD, the reliability of model integration method and platform performance is verified, the time cost of concurrent computation shows that platform has the ability of multiple scenarios calculation concurrently. Hydrodynamic simulation pattern proposed in this paper can provide reference for cloud computing application in scientific computation. The combined communication pattern and model integration method provide technical support for model evaluation and sharing system.

Acknowledgments

This work has been sponsored in part by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China(Grant No. 2013BAB05B03, 2013BAB05B05), Chinese Ministry of Water Resources special funds for scientific research on public causes (201201050), and the State Key Laboratory of Hydroscience and Engineering Tsinghua University (Grant No. 2012-KY-05).We would like to thank all our sponsors and other members of the HydroMP and Cloud Computing Research Group at the State Key Laboratory of Hydroscience and Engineering, Tsinghua University.

REFERENCES

Brooking C. and Hunter J., “Providing online access to hydrological model simulations through interactive geospatial animations”, *Environmental Modelling & Software*, Vol. 43, (2013), pp 163-168.

Burger C. M., Kollet S., Schumacher J., Böselc D., “Introduction of a web service for cloud computing with the integrated hydrologic simulation platform ParFlow”, *Computers & Geosciences*, Vol. 48, (2012), pp 334-336.

Sun A., “Enabling collaborative decision-making in watershed management using Cloud-computing service”, *Environmental Modelling & Software*, Vol. 41, (2013), pp 93-97.

Horsburgh J S, Tarboton D.G., Maidment D.R., Zaslavsky I., A relational model for environmental and water resources data. *Water Resources Research*, Vol. 44, No.5, (2008), W05406.