Study on the Key Technology for Establishing a Cloud Platform-Oriented Digital Oilfield Based on High-performance Computing

Zheng-ping Zhu, Ren-fang Pan[1], Jie MA, Zhe Chen[2], Pang-yun Zhong[3]
[1]College of Earth Science Digital Oilfield Virtual Laboratory, Yangtze University
No.1, Daxue Road, Caidian District, 430100 Hubei, China
[2]Information Center of the Petrochina Dagang Oilfield Company
Dagang District, 300280 Tianjing, China
[3]Information Engineering Institute of RNR Energy Restore Ltd, 12958 Holk Road- Poland
zhuzhengpingzzp@aliyun.com

Journal of Digital
Information Management

ABSTRACT: *This study aims to overcome the principal problems in the current information-support–platform-oriented digital oilfield. These problems include a low utilization efficiency of the software and hardware resources, the high maintenance cost of the system, and limited automatic control. Starting with the existing software and hardware resources and with the current pattern of use of cloud computing, this study designs a cloud platform-oriented digital oilfield, based on high-performance computing (HPC). It also comprehensively describes the architecture of the cloud platform and then solves three key problems, namely, united resource management (computing and license resources), policy-based resource scheduling, and self-service remote cluster visualization. Finally, this paper describes the achievement of establishing and operating a new type of cloud platform-oriented digital oilfield, based on HPC with self-service intelligent cluster development, mainstream application system integration, unified job management, comprehensive real-time monitoring of resources, and statistical statement analysis. The results show that integrating cloud computing into HPC can comprehensively improve the efficiency of use of the hardware and software resources, and reduce system operation and maintenance costs, as well as providing administrators with a friendly and fast way to use and monitor those resources.*

Subject Categories and Descriptors
I.2.10 [Vision and Scene Understanding]: High-performance cloud platform research; **I.4.10 [Image Representation]**

General Terms : High-performance Computing, Cloud Platform

Keywords: Digital oilfield, High-performance Cloud Computing, Resource Scheduling, Intelligent Management and Control, 3D Remote Visualization

Received: 12 July 2014, Revised 19 August 2014, Accepted 25 August 2014

1. Introduction

The goal of high-performance computing (HPC) is to pursue quick and efficient computing performance [1]. The computing performance of the early vectored computers, considered to be the ‘first-generation HPCs’, paralleled computers and computer clusters can mainly be enhanced by improving the CPU frequency. The technical route for this has seen a progression from the traditional single-core to multi-cored processers. So, currently, the running-CPU-cores of HPCs has reached an order of magnitude of 100,000 [2–3]. Therefore, the whole system structure has become more and more complicated, resulting in increases in power consumption and running costs. Meanwhile, the demand for HPC has risen with the continuous development of application research. The traditional technology route is therefore approaching its developmental limits and incurring both high levels of resource wastage and operating costs [4, 5].

Cloud computing is a new service delivery and use pattern, which aims to integrate a group of cheap computing resources through a network to provide a powerful computing capacity. This system is provided to users in the form of a service. The administrators have little need to interact with the service provider yet can configure and use these resources with minimum cost [6–9].

Both HPC and cloud computing consistently provide users with powerful computing capacity [10]. The main difference between them is that HPC solves these problems associated with hardware while cloud computing solves these problems associated management. HPC brings its own problems of management and cost, when their quantity reaches a large scale. On the other hand, cloud computing is used to form integrated computing services through a network based on existing resources and simply requires the solution of the problems of the management of these. In this respect, HPC and cloud computing are complementary. HPC mainly pursues high computing performance, whereas cloud computing mainly aims at high resource utilization [11, 12]. Therefore, cloud computing provides new methods for breaking the development bottleneck of HPC in the application field, by making the inevitable development trend of HPC the 'HPC cloud platform'.

An HPC cloud platform combines existing HPC servers through a network and then provides access to this for users in the form of services. Compared with the traditional HPC platform, the HPC cloud platform can bring at least two advantages. First, it can improve the resource utilization of HPC. Second, it can provide users with convenient and powerful computing services. Aside from meeting the current demands of HPC users, it also can give full play to the unique advantages of cloud computing [13, 14]. At present, the HPC cloud platform is mainly still at the theoretical research stage. Given this background, the author proposes an architectural design scheme for building a cloud platform-oriented digital oilfield, based on HPC, by considering HPC cloud computing (also called equilibrium computing by some scholars [15]) from a theoretical basis, in order to promote the utilization efficient of resources and to reduce the operating cost of those resources.

2. Status of the Digital Oilfield Information - Support-Platform as at the Year 2013

The 'digital oilfield' is an advanced stage and inevitable trend of oilfield information development. It is a technical system that combines the relevant data, resources, software, and knowledge of oilfields with the 'digital Earth' as the guide and the oilfield entity as the object. It is a direct application of the digital Earth in petroleum exploration and production (E&P) [16]. Some scholars believe that the digital oilfield is a logical structure consisting of seven layers, namely, environment (i.e., platform support layer), data, thematic, model, application, integration, and strategy layer [17]. The platform support layer is the core of the logical structure of the digital oilfield, its purpose being to provide environmental support for the whole digital oilfield. HPC is an important part of the information support platform-oriented digital oilfield (ISPODO) [18, 19], mainly used in seismic data processing, seismic interpretation, reservoir simulation, reservoir description and other core research areas in (E&P).

HPC is also an important branch of computer science and is primarily used in fields that require large-scale scientific computing. After decades of development, HPC has become the principal mainstream hardware structure of the ISPODO. As an illustration, consider the information support platform of an oilfield branch of the China National Petroleum Corporation (CNPC) which has been configured with 54 high-performance blade servers, of which 48 support the conventional interpretation and reservoir simulation and 8 support 3D interpretation software. Its storage capacity has reached 130 TB, while the overall HPC environment has reached 128 CPUs and 1024 cores, with the total memory expanded to 8 TB. The platform can also support the daily research work of over 200 users. On this basis, the application software has been installed to correspond with the need to realize the centralized development, application, and management of the infrastructure and software through the team work space (TWS) portal system and the independent emulation facilities, which forms the pattern of the current HPC ISPODO (Figure 1).

An in-depth study of the core business of the digital oilfield shows that the oilfield has gradually entered the middle and late periods of its development. The fractured, thin, subtle, and special lithology reservoirs which are under development mainly exist in complex geological environments; these reservoirs are buried deeper, are more subtle and more difficult to explore[20]. The current ISPODO is also increasingly being exposed to new problems which include:

(1) Poor utilization of software and hardware resources: The demand of the digital oilfield for computing resources and licenses is not always static but is often unexpected. The demands from researchers for computing resources in a particular period might suddenly increase above the average running demand. However, in the frame of the existing information systems (Figure 1), all professional application software is firmly installed on the server nodes. If the IT department handles the peak load through purchasing clusters, this method becomes extremely expensive. If not, the business research demands for computing resources may not be met.

(2) High system maintenance costs: The large-scale use of specialized professional application software in the business research oriented digital oilfield means a complexity of structure, and maintenance difficulties, with resulting extremely high costs. Furthermore, many types of professional application software are required (Figure 1). The system architecture, operating system, and demands on hardware resources of each type of professional application software are also different, thereby aggravating the maintenance cost of the system.

(3) Little automatic control: The traditional HPC model lacks real-time monitoring or alarms for abnormalities in the operating status of the system resources (licenses, databases, servers, switches, storage etc.), so this will obviously affect the operational efficiency and the emergency response mechanisms of the system [21].

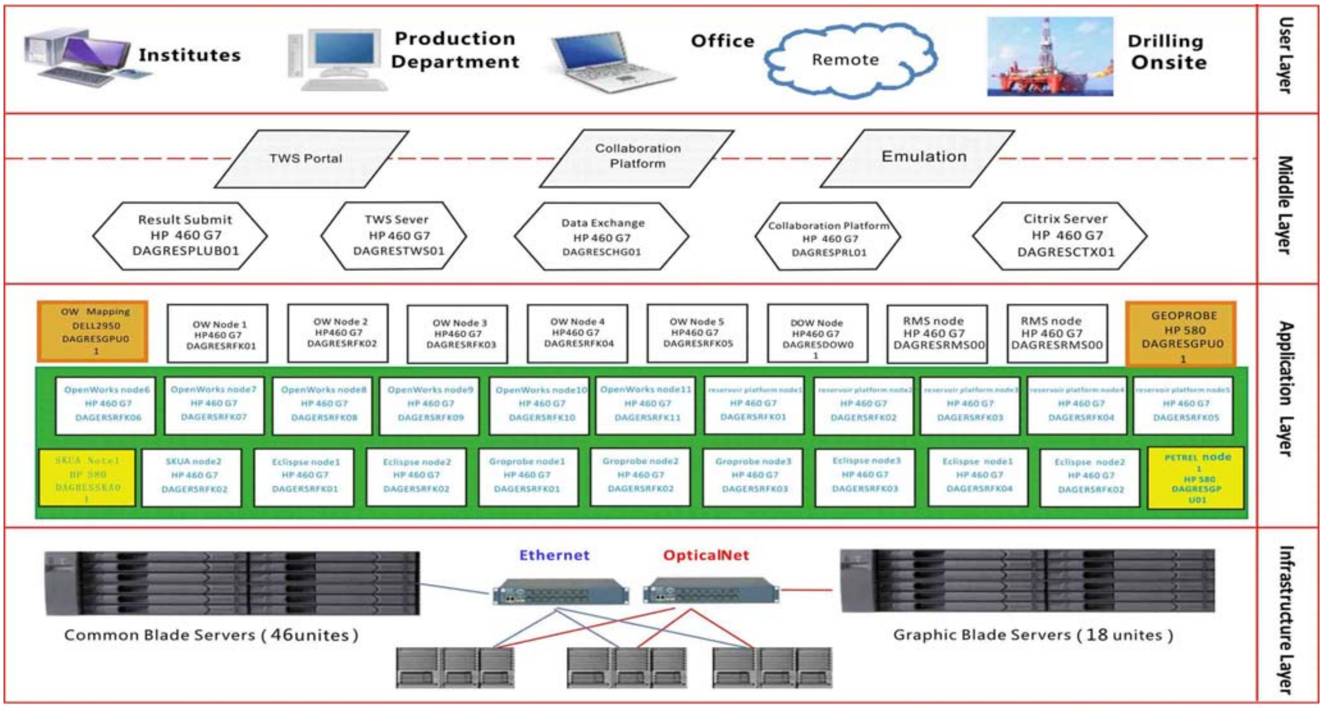


Figure 1. Overview of HPC ISPODO As at The Year 2013

3. Functional Design of HPC Cloud Platform-Oriented Digital Oilfield

The most effective solution is to break the traditional model of tightly coupled nodes and build a system of loosely-coupled cloud platform nodes to meet the above-mentioned three principal challenges. The key points of such a solution are mainly focused on: a scalable, elastic computing and storage configuration, efficient resource scheduling, remote visualization of the graphic servers, flexible cluster allocation, and efficient resource monitoring to improve the platform efficiency and to reduce operational costs.

Therefore, based on the advanced concept of HPC cloud computing, a HPC cloud platform-oriented digital oilfield (CPODO) can be designed and established (Figure 2). Through the integration of the existing infrastructure, the platform can provide the company with a convenient working environment, and its employees can conduct seismic data processing, seismic interpretation, and reservoir simulation without going into the computer room where the servers are located, since they can obtain all the required resources to conduct their daily research simply at the click of a mouse by logging into the collaborative research portal.

As shown in Figure 2, from the perspective of the cloud service model the HPC CPODO can be divided into an infrastructure service (IAAS), a platform service (PAAS), and a software service (SAAS), and into an infrastructure layer, a virtual pooling layer, a middleware layer, and a user layer from the point of view of its functional structure. Through the HPC-based cloud management platform (PAAS), the computing, storage, and other resources can be rationally divided to provide a unified access interface for all the different users. Then, users can log into the united portal interface and select the relevant application software which they need to access. At the same time, administrators can also conduct monitoring, statistical statement analysis, and management of the running of the platform. The key components are as follows:

(1) Cluster resource pooling: All the computer resources (physical servers, virtual servers, graphic workstations, storage, networks etc.) are abstracted into one unit to provide users with pooled computing, storage, and network resources through a unified system of organization and management.

(2) Automatic configuration services: The clusters are set up according to the different demands of each application system for hardware resources. These would typically include the interpretation, modeling and simulation, 3D interpretation, and engineering-design clusters. Then, the relevant computing resources (physical or virtual computing nodes) and storage are allocated to the corresponding cluster. Finally, the application software can be installed for each relevant application cluster. Moreover, all the software only needs to be installed once, as it can be automatically installed in future, according to demand.

(3) Resource intelligent scheduling and distribution: On the basis of realistic demands determined through studies of the business and the operating characteristics of each application system, an intelligent resource-scheduling strategy can be custom-made to achieve highly efficient use of the cloud resources. and developed through a convenient cluster design template.

(4) Cluster Visualization Template Design: Personalized clusters for various application systems can be designed and developed through a convenient cluster design template.

(5) Resource Running Real-time Monitoring and Statistical Statement Analysis: The real-time monitoring and statement analysis function can monitor the resource running efficiency and tendencies, in real time, in order to provide a good reference for the design of future resource expansion and configuration schemes.

4. Key Technology

4.1 Unified resource management

The core objective of cloud computing is to integrate the IT resources that are connected by a network, to constitute a resource pool so that these can responsively meet the increasing computing demands of users. Therefore, the hardware infrastructure and software resources need to be abstracted into a single unit to achieve unified resource management [22, 23]. To meet the requirements of the CPODO, the main factors required to realize such united resource management are as follows:

4.1.1 Computing Resource Management

Virtual technology is an important means, although not the only one, for ensuring the unified management of the computing resources. Particularly for high-performance servers, virtualization is bound to consume a certain amount of the computing resources. Therefore, according to the different needs of the application software for the operating system, the computing resources can be divided into two main types: the physical and the virtual servers. Meanwhile, respective configuration templates are set up for these. The administrators can then set up the relevant computing node by selecting a template, and adding it to the relevant computing cluster, in order to realize the unified management of the computing resources (physical and virtual servers).

The physical sever template is principally aimed at Linux-based application software. The standardized template of the physical server mainly includes the Linux operating system, the NIS domain, the storage, design, network configuration of the mount point, the server name and the computing clusters. The virtual server template provides for Windows-based applications, with its template based on a physical server, deployed in mirroring mode. Customization of the template then mainly includes creating a virtual server, installing a Windows-7 operating system, the application software, and the computing clusters.

During the initial period of use, the computer resource can be run in a traditional fixed system resource allocation approach. After this initial period, appropriate strategic adjustments can be conducted at any time, according to the statistics revealed in the monitoring reports. In aforesaid example with 54 blade nodes in the ISPODO, the servers are primarily distributed according to their functions: management nodes (2 sets), seismic interpretation (15 sets), velocity analysis (11 sets), reservoir simulation (4 sets), structure modeling (8 sets), geological modeling (graphic workstations, 8 sets), reservoir prediction (2 sets), and load monitoring (6 sets). These nodes can be configured according to the following script files:

```
Begin HostGroup  // the configuration initialization Velocity analysis
(dagresrfk[01] dagresrfk0[5-9] dagresrfk10 dagresrfk1[4-7])
 Reservoir simulation
(dagresrfk18 dagresrfk19 dagresrfk20 dagresrfk21)
 Structure modeling    (dagresrfk22 dagresrfk2[4-10])
 Reservoir prediction   (dgresptr04 dgresptr06)
 Geology modeling    (dagrespet0[1-8])
 RTM_ONLY
 (dagresora[01] dagresora[02] dagresrfk[01] dagresrfk[04]
dagresrfk[05] dagrespmb[01])
 Seismic interpretation (dagresrek0[2-9] dagresrek1[0-6])
 End HostGroup //end of the configuration
```

4.1.2 License management

The cost of the software license resource is significantly high. In a typical situation, enterprises are unlikely to choose to purchase enough license resources to accommodate the peak demands from users. Therefore, lack of availability is sometimes inevitable, as a result of the insufficient number licenses. With the license management module provided by a cloud platform approach, the license resources can be merged into the policy-scheduling, which can be dispatched in the same way as the CPUs. Thus, current problems can be solved effectively. For example, licenses can be issued on the basis of priorities, and can be temporarily seized for high-priority tasks, while the work within the robbed application is suspended. The licenses can be used by departments according to a ratio designed to realize fair distribution, so the long-term vicious-circle of constraints because of limited numbers of licenses can be terminated.

Oilfield enterprises often adopt a ‘project-team and cluster’ configuration. This strategy for license distribution can be customized through such an arrangement to share license resources among departments, to ensure that the license-sharing by ratio also achieves a highly efficient use of the licensed resources.

Based on the cloud platform, oilfield enterprises usually set up high-performance clusters for particular application systems and then define many project teams within the clusters in order to conduct the license-scheduling, by regarding each project team as a unit. Taking reservoir modeling software as an example, a typical configuration script would be as follows:

Begin ServiceDomain // the configuration initialization
NAME=RersMod_Server // the name of the license server
LIC_SERVERS=((1888@hostD)(1888@hostE))
//the license server host name
DISTRIBUTION = RersMod _Server (Project_cy 2/1 Project_kt 1)
//Set the project permits allocation proportion
End ServiceDomain // End of the configuration

Based on the above allocation strategy, we can define the scheduling of reservoir modeling software licenses with a production-team (Project_cy) and exploration-team (Project_kt) as two separate project teams, which use a common reservoir of modeling software licenses. The ratio of the two projects is 2:1 respectively. When the production-team's licenses are not in full use, the exploration-team may borrow a certain percentage of them. When the production-team is in need of its licenses at a later point, its request for the licenses is immediately granted so that those licenses being used by the exploration team are returned.

Based on project-team-scheduling, the licenses can be allocated within a cluster, while, based on cluster-scheduling, the licenses can be allocated among the clusters according to specified proportions. From this point of view, licenses can not only be allocated within each cluster but also among different clusters. Compared with project-team-scheduling, cluster control provides more flexible scheduling and more efficient distribution, in that, there is an addition of the corresponding configuration parameters on the basis of the above scheduling strategy based on the project-team. The configuration script information for a cluster-scheduling approach would be as follows:

CLUSTER_MODE = Y //defines whether to allow between clusters deployment, Y (yes), N (no)
CLUSTER_DISTRIBUTION = RersMod Server (Cluster1 1 Cluster2 1) // Set the mixing ratio between the cluster

4.2 Policy-based resource scheduling

A resource scheduling strategy provides users with abundant scheduling strategies to meet their computing demands from the system, according to changing platform use. The strategy mainly includes resource distribution, resource placement, resource removal, and resource dynamic adjustment strategies [24].

Considering the key characteristics of the ISPODO, this platform needs to be focused on the following three aspects: high-efficiency of utilization of the computing resources, reducing the vacancy rate of the high-performance graphic servers, and resourcing the needs of any urgent computing task in a specific time period. For this purpose, the following three strategies are mainly adopted to distribute the computing resources:

(1) Fair scheduling: The computing resources are fairly and rationally used according to the distribution ratios that the administrator offers to the users or user groups.

(2) Dynamic scheduling (over time): The high-performance graphic servers provide an operating environment for the 3D visualization interpretation software. However, users will probably not be operating their graphic servers overnight or during holidays, thus freeing them up completely during these periods. Therefore, a time-based configuration strategy can be set up optimizing the availability of such high-performance graphic servers.

(3) Preemptive scheduling: A preemptive scheduling level can be defined to allow the instruction system to decide the work-order according to the appropriate levels needed to execute the scheduling.

Based on the features of a variety of professional application software systems, the expansion or shrinkage of the configuration scripts for this software can be designed to achieve automatic changes for their use of the computing resources on the cloud platform. The script below is an illustration of the configuration of the reservoir modeling software, on the basis of the aforementioned three resource scheduling strategies:

Begin Queue // initialize the scheduling
QUEUE_NAME = RersMod_q // the scheduling queue name
PRIORITY = 30 // the scheduling queue priority level
USERS = ecl2013 // the scheduling queue user name
FAIRSHARE = USER_SHARES[[cy1c, 2] [cy2c, 2] [kt, 1] [others, 1] //according to the share of the fair scheduling, set the fair scheduling policy // as the time changes, set the figure station to compute the resource scheduling strategy
if time (7:00–20:00); HOSTS = RersMod_Server
//7 a.m. to 8 p.m. during day period, then schedule reservoir modeling computing resources
else HOSTS = RersMod_Server; endif
// during the other time, automatically figure station computing resources assigned to reservoir modeling
PREEMPTION=PREEMPTIVE
//preemptive scheduling strategy setting can preempt the lower precedence than the queue resources
PREEMPTION=PREEMPTABLE
//preemptive scheduling strategy setting can be high-priority queue task preemption
End Queue // the scheduling ends

4.3 Self-service cluster remote visualization

Most application software in the digital oilfield system frame have to conduct 3D modeling research and then rotate, slice, and render the models. Some even use GPU for the model calculations. These operations require a high-grade GPU graphics server [6]. However, when multiple users share a server, the GPU graphics server must be able simultaneously to support those multiple users; otherwise, the expensive graphics server cannot be utilized efficiently. With the increasing scale of such enterprises, the unified management of graphics server resources, to achieve remote visualization for user access, has become the inevitable development trend.

Therefore, the (CPODO) can incorporate a remote 3D visualization scheme for the self-service cluster (Figure 3

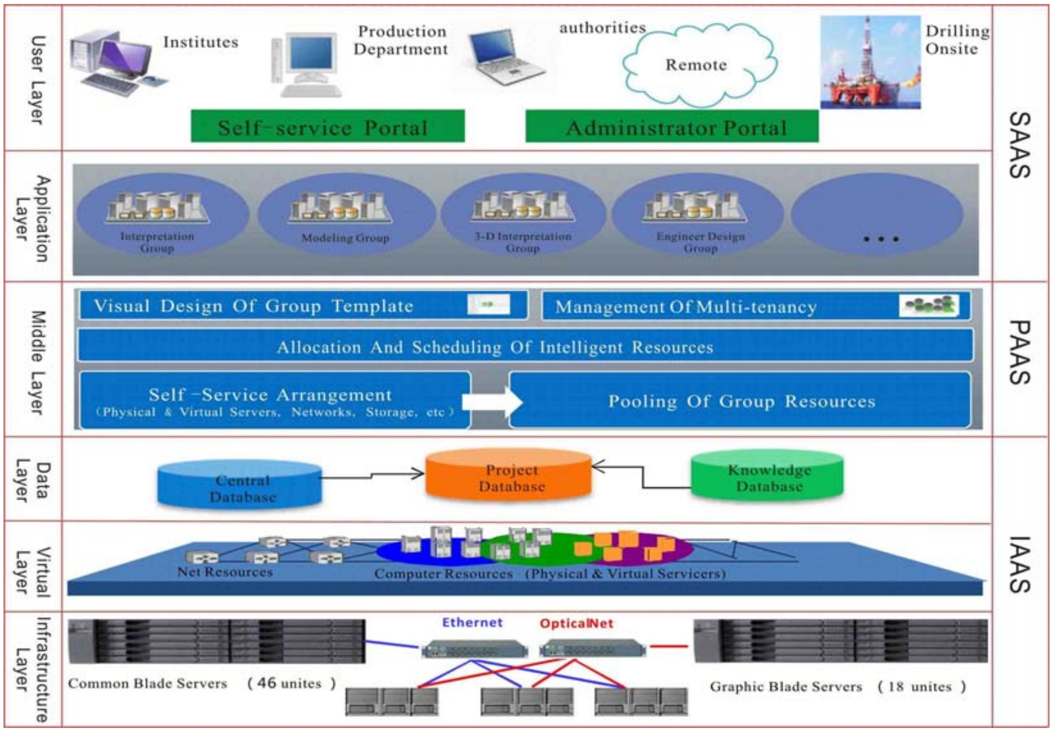


Figure 2. Overview of The HPC CPODO

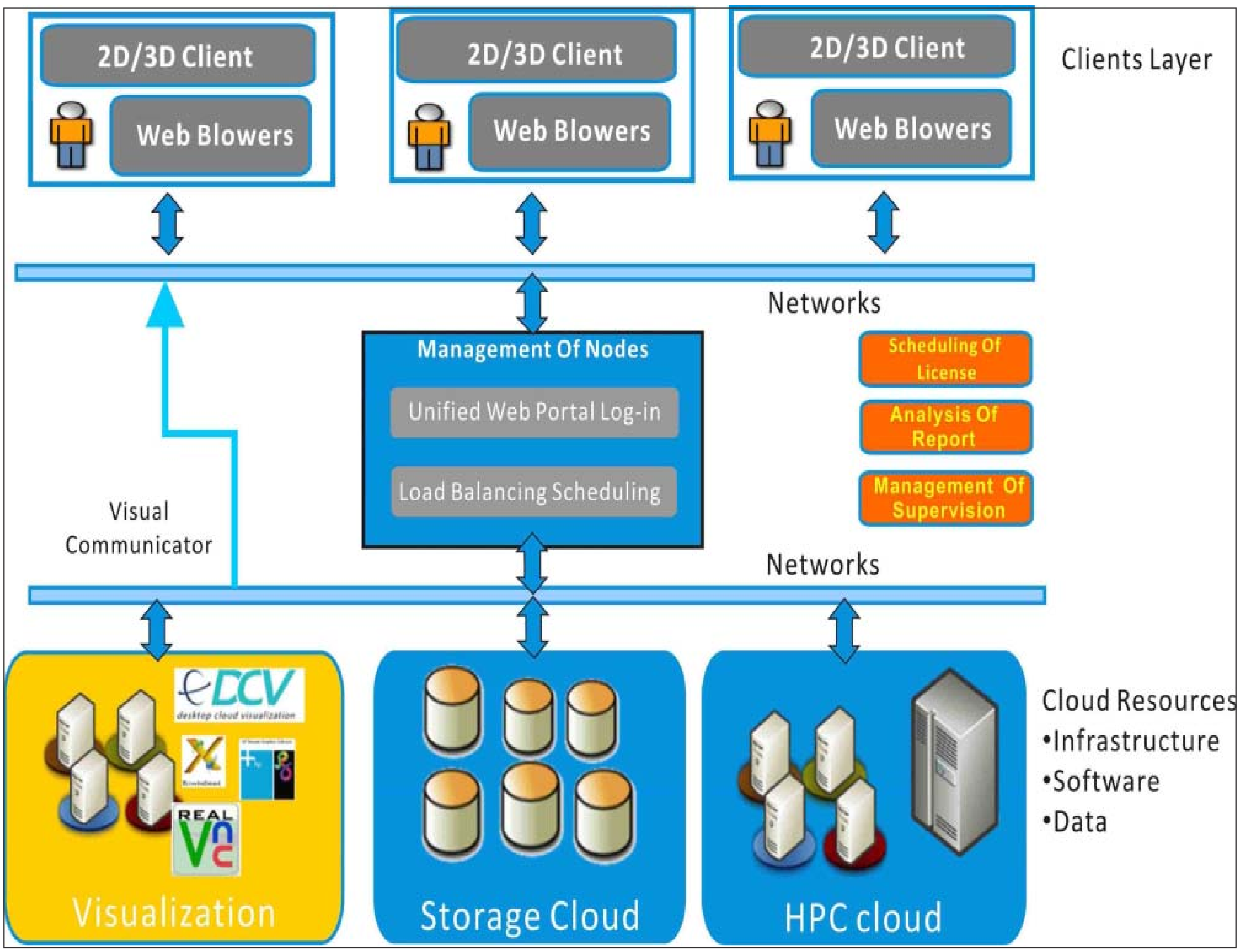


Figure 3. Overview of Self-service Cluster Remote Visualization Scheme

). The following are the main characteristics of the scheme:

4.3.1 Self-service

To utilize the graphic resources of the cloud platform, users only need to log into the cloud platform and submit an application for 3D visualization services. The cloud platform automatically assigns the graphics resources to the users. Resource application and approval are guided by using the cloud template design. During the utilization period, the cloud platform automatically recycles resources without the need for administrator or user intervention. Recycled resources continue to enter the pool for user applications. When the user application is successful, the system assigns a login IP address, login user name, and password to the user by email. When the term expires, the system sends a notification email, so the users can choose to recover or to extend those resources.

4.3.2 Cluster

The visualization system adopts cluster scheduling by the cloud management platform, which includes user Web access, resource scheduling of the graphics server, license scheduling, job submission, real-time monitoring, and statistical analysis. The user can log into the Web portal and select the software from the software list. Then, a software running request for license resource scheduling is sent to the scheduling server. When the license request is met, then the scheduling server begins to check from the graphics server to determine whether there are sufficient resources to run the software. If the current resources are sufficient, the server allocates an appropriate graphic server to the user.

4.3.3 Remote

In traditional 3D visualization, the application software sends the graphics process command to the graphics server through a local high-speed I/O channel, and then the graphics server paints these commands to images to display on the local screen. However, in this study, the remote visualization can be achieved with desktop cloud visualization (DCV) [25, 26]. After the user applies for use of the graphics server, the allocated graphics server is prompted, and then all the data operations and graphics painting run on the remote graphics server. Finally, the image on the graphics server is sent to the terminal desktop through the DCV website agreement.

5. Analysis of Application Effect

The HPC CPODO realizes self-service intelligent cluster allocation, mainstream application system integration, united job management, real-time monitoring of the overall resources, and statistical statement analysis, which improves the utilization of software and hardware resources and reduces the operational costs. This outcome is achieved by using existing computing, storage, and network resources. Thus, the main problems and challenges in the ISPODO are efficiently overcome. The HPC CPODO is also regarded as the first comprehensive cloud environment for petroleum industry research in China.

5.1 Self-service intelligent cluster deployment

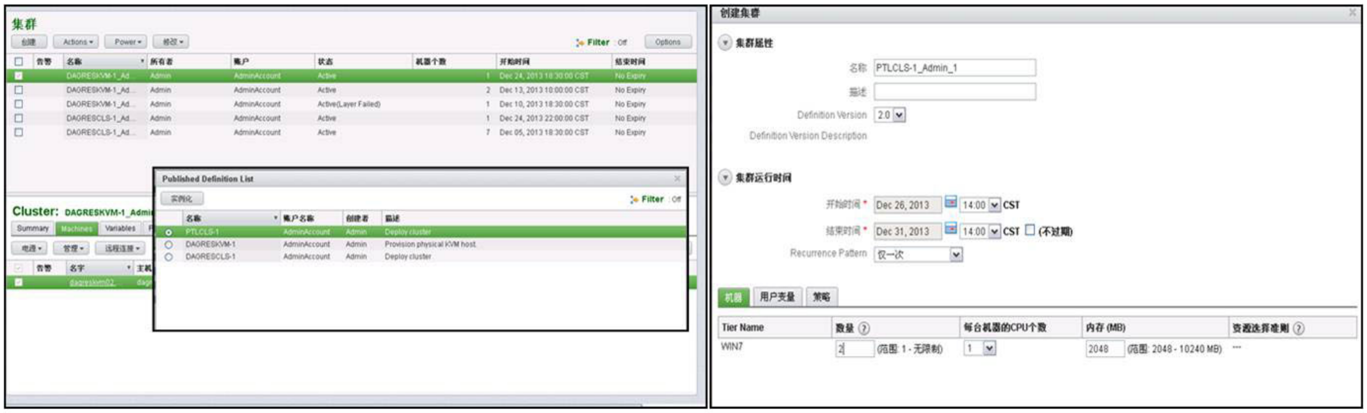
This platform provides a simple and friendly self-service portal interface. The user only needs to log into the self-service portal platform as required by the unified authentication function module provided by the portal which then makes intelligent customization decisions, based on demand. The user is then allocated computing resources within the platform, while quick configuration of the physical servers or virtual servers is achieved using the predefined templates, and finally he is added to the cluster. A set-up workload which would previously have taken one hour can therefore be completed within 5 minutes.

On the basis of the predefined templates of the physical server and the virtual server mentioned in 4.1, the user can log into the cluster self-service definition interface (Figure 4a), select setup (or modification) on the upper right corner of the interface, and then select the corresponding physical server or virtual server template on the pop-up cluster selection interface, to achieve rapid setup and dynamic adjustment of the computing cluster. Taking the expansion of the geological modeling cluster as an example, the user can select the virtual server template of PTLCLS-1; thereafter, the user can input the corresponding configuration parameters, such as cluster name, description, cluster running time, and number of the allocated server, on the pop-up template instantiation parameter settings interface (Figure 4b). Finally, the user clicks on "*Create*" for the platform to start, automatically, to allocate the virtual server PTLCLS-1 on the corresponding computing node and automatically distribute the virtual server to the geological modeling computing cluster. The entire process requires no further user intervention. At the same time, all the allocated resources can be directly delivered for use. The application clusters that have been configured can also be viewed modified, or terminated through the self-service portal website.

5.2 Mainstream application system integration

The platform replaces the traditional decentralized management "*trinity*" pattern (a researcher, a workstation, and professional software) but can uniformly install the seismic interpretation, geological modeling, numerical simulation, engineering design, and other professional software. Users can conveniently select the required professional application software simply by logging into the united portal interface. Thus, the users are provided with a comprehensive application software service.

As shown in Figure 5, the application software has been successfully centralized and installed for use. Users can immediately enter the work submission interface with just a click on the mouse. Both the high-performance servers and the remote 3D visualization servers can be accessed from the portal application center. The cloud pattern can



(a) Cluster Self-service Deployment Definition Interface (b) Template Instantiation Parameter Settings interface

Figure 4. Self-service Intelligent Cluster Definition Interface of the HPC CPODO

Figure 5. Login Page of Mainstream Professional Application Software Integration Centre

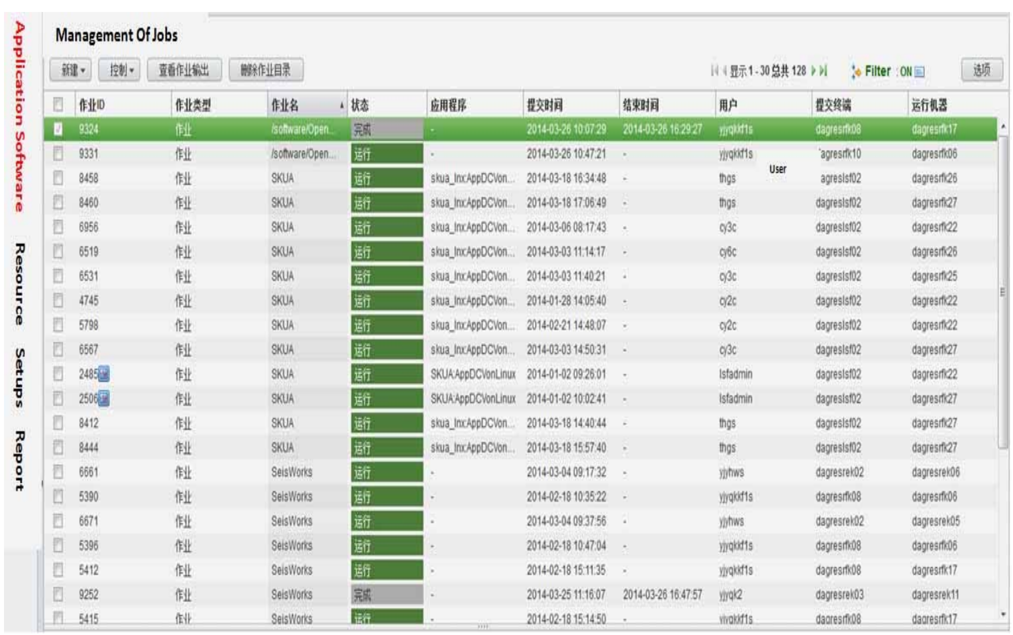


Figure 6. Unified Job Management Portal Page

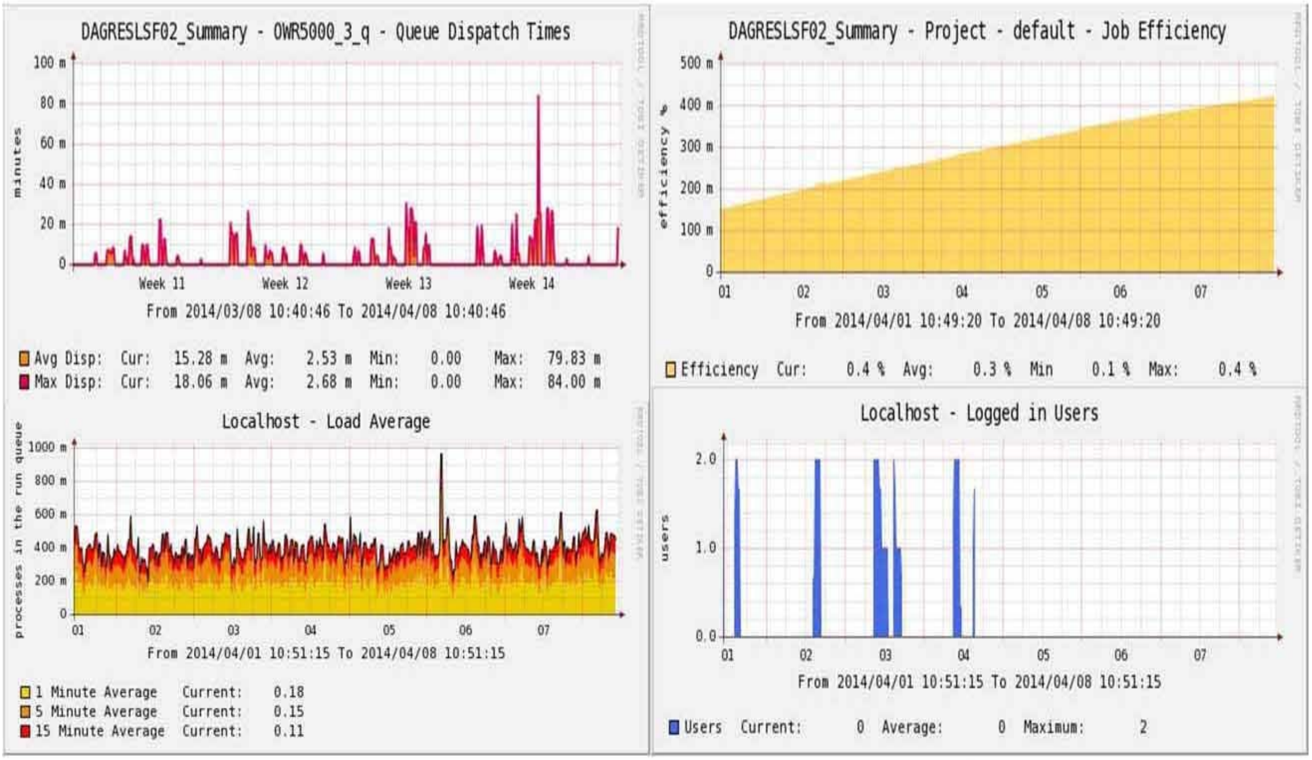


Figure 7. Resource Monitoring and Statistical Analysis

provide users with professional application software without the need for maintenance and management of system, in order to let users to focus on the business research and software usage.

5.3 Unified job management

On submitting a job, the user can see the job's running status on the portal page at any time, and check the data and output information, which are generated in the calculation process (Figure 6). Each row shows the job name, running state, application program, submission time, completion time, user name, running node, and other information for a particular job. This feature allows the user conveniently to control the running status of all the jobs submitted. Even job running data submitted several months before can also be reviewed.

The cloud platform administrator can review all the users' work, therefore being fully aware of their usage statistics. Any abnormal work can be handled on the portal interface without needing to seek out the running node and log into it. Thus, the running efficiency and stability of the system are improved significantly.

5.4 Comprehensive resource real-time monitoring and statistical statement analysis

This platform provides a set of monitoring tools that can comprehensively monitor all the computing, management, and database nodes on it. On the monitoring interface, the user can easily see the node load situation, including CPU, memory, and their running status The comprehensive data status of the cluster, the usage status of the users, the running status of all assignments, and the resources consumed for the work can also be viewed. Aside from the monitoring function, the platform can provide a rich statement analysis function by periodically drawing off statistics showing the information usage and offering inquiries for defined periods. Therefore, the use and distribution information for all the licensed services over a period, the utilization of the licenses, and the user distribution can be shown. Through such data analysis over the entire year, the administrators can understand the general trends of license use, thereby providing a significant reference basis for decisions on whether or not to increase the quantity of licenses in the following year while to avoiding waste resulting from blind purchases. From the cluster load tendency statement (Figure 7), the administrators can learn the usage trends of the computing resources of the entire platform, and, through analysis, can assess whether the existing hardware environment can meet the current demands, predict which period will require peak usage of the computing resources, and conduct flexible adjustment of these resources in this phase, ahead of such time. In this way, the platform can help users to enhance the productivity of the oilfield, by providing an important reference from which to formulate scheduling strategies.

6. Conclusion

The findings of our study can be summarized as follows: (1) The integration of the cloud computing concept into HPC can realize unified pooled-resource management and provide a strong heterogeneous platform management ability that can support the physical server, virtual server, and Linux and Windows operating systems as well as provide for self-service intelligent cluster allocation. The integration improves resource utilization and reduces system maintenance costs.

(2) The policy-based resource scheduling not only provides reliable support for HPC and simplifies the work load of the IT administrators but also ensures efficient resource scheduling of the licenses, graphics servers, CPUs, etc., which improves their utilization.

(3) The unified portal interface that the cloud platform provides has achieved centralized installation with one-key login to mainstream professional application software, so that users can quickly find the software resources that they need, submit jobs, and view the results in real time.

(4) The monitoring platform displays the usage across the entire platform clearly at a glance. Many statistical statements can also reflect the latitude data of the platform, providing a favorable basis for the information center to manage clusters and to establish scheduling strategies.

References

[1] Pang, Shi-ming. (2002). Application of High Performance Computing Technology in Oil-Gas Exploration and Production. *Progress in Geophysics*, 25 (1) 9-12.

[2] Feng, W. C. (2006). The Importance of Being Low Power in High-Performance Computing. *CT Watch Quarterly*, 1 (3) 11-20.

[3] Gordon Bell., Jim Gray. (2002). What's Next in High-performance Computing. *Communications of the ACM,* 45 (2) 91-95.

[4] Chung-hsing Hsu., Wu-chun Feng. (2005). A Power-Aware Run-Time System for High-Performance Computing. Proceedings of ACM/IEEE Conference on High-Performance Computing, pages 1-9.

[5] Jun-hua Zhang., Sheng-tao Zang., Lian-yu Shan., et al. (2010). Current Development Status and Tendency of High Performance Computing. *Petroleum Geophysical Exploration,* 45 (6) 918-925.

[6] Peter Mell., Timothy Grance. (2009). The NIST Definition of Cloud Computing. National Institute of Standards and Technology, 53 (6) 50-56.

[7] Michael Armbrust., Armando Fox., et al. (2010). A View of Cloud Computing. *Communications of the ACM*, 53 (4) 50-58.

[8] Luis, M., Vaquero., Luis Rodero-Merino., et al. (2008). A Break in the Clouds:Towards a Cloud Deûnition. *ACM SIGCOMM Computer Communication Review*, 39 (1) 50-55.

[9] Fang, Lei. (2011). Key Technology Exploration and Research of the Land Resources Services High Efficient Disposal Platform based on Cloud Computing. Zhejiang University, p 55-80

[10] Vecchiola, Christian., Pandey, Suraj., Buyya, Rajkumar . (2009). High-Performance Cloud Computing:A View of Scientific Applications. Pervasive Systems Algorithms and Networks (ISPAN) 10th International Symposium on IEEE, p 4-16.

[11] Mark, F., Mergen., Volkmar, Uhlig., et al. (2006). A case for High Performance Computing with Virtual Machines: Proceedings of the 20th Annual International Conference on Supercomputing ACM, p 125-134.

[13] Youself, L., Butrico, M., Silva, D. D. (2009). Toward a Unified Ontology of Cloud Computing. *In*: Proc of Grid Computing Environment Workshop, p 1-10.

[14] Keith, R., Jackson., Lavanya Ramakrishnan., et al. (2010). Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud. Cloud Computing Technology and Science Second International Conference on IEEE, p 159-168.

[15] Na, Jiang. (2011). Intel HPC: Equilibrium Concept and Petroleum Exploration Ten Billion Computing. *Petroleum & Equipment*, 2011 (4) 76-77.

[16] Chen, Qiang., Hong-lin Wang. (2002). Data, Information, Software and Knowledge of the Digital Oilfield and Integrated Oilfield. *Petroleum Geophysical Exploration*, 37 (1) 90-96.

[17] Wang, Quan . (2003). Digital Oilfield Mode and Development Strategy Research. Tianjin, Tianjin University, p 43-47

[18] Zhao, Gai-shan., Bao, Hong-lin. (2001). Application of Cluster Computing Technology in Oil Industry. *Petroleum Geophysical Exploration*, 40 (3)118-126.

[19] Li, Ming. (2014). Development of High Performance Parallel Computer and Its Application in Oil Exploration. *Science & Technology Review*, 32 (2) 80-83.

[20] Tian, Zuo-ji., Wu, Yi-ping., et al. (2014). Appraisal and Potentiality Analysis of Global Regular Oil Resources. *Geoscience Frontiers*, 21 (3) 10-17.

[21] Wang, Jing-de., Li., Jian-yu., Fu, Xi-chun. (2013). Realization of the Integrated Dynamics Monitoring Software of Multi-clusters. *Information Technology,* 37 (10) 106-109.

[22] Calyam, P., Rajagopalan, S., Seetharam, S., et al. (2014). VDC-Analyst: Design and Verification of Virtual Desktop Cloud Resource Allocations. Computer Networks, p 1-10.

[23] Tao, Xie. (2012). Set up Private Cloud Platform to Improve IT Resource Utilization of China Petroleum. *Software and Information Service*, 10 (10) 54-54.

[24] Sun, Rui-feng., Zheng-wen Zhao. (2010). Resource Scheduling Strategy based on Cloud Computing. *Aeronautical Computing Technique*, 40 (3) 103-105.

[25] Torterolo, L., Ruffino, F. (2011). Desktop Cloud Visualization: The New Technology to Remote Access 3D Interactive Applications in The Cloud. Studies in Health Technology and Informatics, 2011 (175) 111-112.

[26] Torterolo, L., Papaleo, G., Scaglione, S., et a. (2012).

3D Cloud in Life Sciences: An Innovative Framework for Remote 2D/3D Visualization and Collaboration. Computer Based Medical Systems (CBMS) 25th International Conference on IEEE, p 159-168.