

Design and Implementation of Knowledge based Cloud Platform for NSFC Service System

Jianjun Li¹, Yuhui Zheng² and Jin Wang²

¹ *Information Center, National Natural Science Foundation of China,
Beijing 100085, China
E-mail: lij@nsfc.gov.cn*

² *College of Computer & Software, Nanjing University of Information Science &
Technology, Nanjing 210044, China*

Abstract

With the quickly development of Natural Science Foundation of China (NSFC) foundations, management of mass fund-relevant digital data, such as researchers and institutions, academic achievement, funds and so on, is becoming more and more important for supporting basic scientific research. In this paper, we fully introduce the design and implementation of NSFC cloud platform that will be implemented in the service system of knowledge base. Firstly, the whole cloud platform architecture is presented. Then we introduce the system software and equipment environment of our platform. Lastly, we present the construction of private cloud, network components, storage components, and computing components, respectively, with illustrative figures and tables.

Keywords: *Cloud Platform; Mass Storage; NSFC; Information Sharing*

1. Introduction

As a new computing model, Cloud computing [1-3] can share IT resources through virtualization based on networks as Internet. Its core idea is achieving resource integration and configuration optimization by unitized management and scheduling of computing, storage, network, software and other resources, for satisfying various demands from different users, such as access and extension services readily, low-cost and so on. The technique has been considered to be a significant reform in information technology after personal computer and Internet. And it will bring about fundamental change in the way of working and business models. With the rapid growth of global markets of the cloud computing, the countries and areas from all of world have paid high attention to cloud computing, and view it as the key area of future strategic industry. Many of developing strategies and plans were formulated to speed up the deployment of cloud computing infrastructure at the national level, with the purpose of seizing the cloud computing industry commanding height.

In February 14th of 1986, the State Council approved the establishment of the National Natural Science Foundation of china (NSFC). After many years of development, with firmly insisting on supporting basic research, the NSFC has developed and formed a stable funding structures composed of the three series of components: research foundation, talent foundation and environmental supporting foundation. NSFC's responsibility is to use the granting funds effectively in line with the developing policy and planning of national science and technology, to support basic research, persevere in free exploration, produce a guide effect, cultivate talents, and advance the development of science and technology. In the process of long-term basic research investment, the NSFC has accumulated a lot of management date on basic research, including structured and unstructured data of researcher, researching results, research activity, researching

environment and funding management. Based on the current existing core data resource, sustainably using advanced information technology to develop the national basic research knowledge service system has important and far-reaching significance.

The national basic research knowledge service system is not only a platform for the National Natural Science Foundation of China, but also a platform to serve the majority of scientific research personnel. It is not only a platform for protecting the intellectual property rights of scientific research achievements, but also a platform of sharing the resources and achievements of scientific research institutes. In addition, it is platform to show our latest achievements in scientific research and promote scientific research information disclosure in our country, but also a platform to realize long-term preservation of Chinese scientific research resource, unified management and sustainable utilization.

The above-mentioned system, shown in Figure 1, mainly targets the basic research personnel, fund management personnel, research policy makers and the community. Additionally, it focuses on the field of basic scientific research, gathers all types of data resources including projects, results, news and so on, from multi-source heterogeneous data sources (such as resources of fund, professional databases, Internet, etc.) to generate the big data of basic research. Moreover, it can extract information, discover knowledge from big data, and build basic scientific research repository, with automatic analysis technology or semiautomatic analysis technology combining the expert wisdom and the collective wisdom. On that basis, it also can provide various professional knowledge services for different kinds of users, for instance in-depth search, multi-dimensional navigation, intelligent recommendation, thematic generation and so on, by supporting the upper application using knowledge from the repository. Based on this, through taking advantages of knowledge in the repository to support the upper application, the system can provides in-depth search, multi-dimensional navigation, intelligent recommendation, thematic generation and other professional knowledge services for various users.

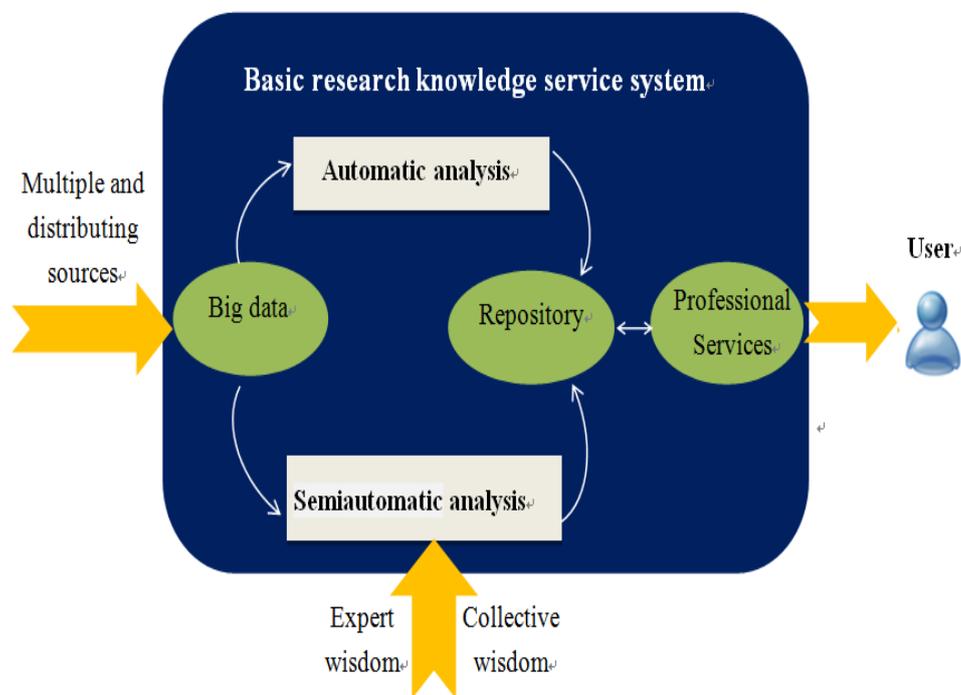


Figure 1. Schematic Diagram of Basic Scientific Knowledge Service System

To date, NSFC has accumulated massive academic and management data from the projects, experts, institutions and achievements (for example research papers, dissertations, technical reports, patents, awards and etc.) and varieties of project reports. Based on some statistical reports, about one million of project applications were proposed and 160 thousand of the projects were granted approximately. In addition, there are 2 million participants, 3200 scientific research affiliations and 1.6 million electronic achievements in the NSFC data repository[4,5]. With the quick growth of massive data, it is urgent to make full use of cloud computing, cloud storage techniques to construct next generation explore new basic platform.

With this aim, we introduce a novel program for cloud platform construction to support basic research knowledge service system construction of NSFC in this paper. It mainly contains the requirements of platform construction and the specific implementation of our proposed program. The latter can be further divided into the room environment construction, the private cloud structure, and the system software configuration and so on.

2. Related Platform System

We have researched and analyzed the domestic and foreign similar websites. The primary results are as follows: 1) Online grants management for The National Science Foundation (NSF, United States) (www.research.gov); 2) The European Commission (EC) —research & innovation official website (www.ec.europa.eu/research/index.cfm); 3) Google's scientific research official website (United States) (www.research.google.com); 4) Microsoft's research and academic official website (United States) (www.research.microsoft.com); 5) IBM's research and academic website (United States), (www.research.ibm.com)

2.1. Resources Classification

Comprehensive analysis on the above mentioned scientific research websites shows that, their resource and content classification are similar as a whole. We mainly consulted NSF and research & innovation official website, because their functions and orientation are closest to our NSFC sharing service network. Most of the resources therein are text-based resource, including some downloadable PDF files. Moreover, lots of pictures and few video documents are included.

2.2. Research Information

1) Achievements. According to the scientific field and geographic information, these platforms mainly contain published journals and downloadable papers.

2) Personal information. Binding with achievement and project information, personal information mainly include a series of data about various participants, such as researchers, scientific research institution, groups and investor, who engages research activities.

3) Project. Project information mainly include research field, funding, situation of application, relevant personal and achievement information etc.

4) Geographical information. The classified resource can be queried according to different geographical area information. The user can read the related scientific research information by clicking the electronic map. Moreover, the statistical number of projects, achievements of scientific research can be provided.

5) Hot research Topics. The popular research fields and methodologies which attract great focus and funds have been listed.

6) Project achievement reports. Research director or cooperator provides a brief introduction or abstract related to the projects, and corresponding academic achievements.

7) Academic activities. The lists of recent academic conferences and activities hold by various universities are provided.

2.3. Resource Display Methods

As is known that there have been massive amount of academic achievements in the NSFC service system. Most users need to exactly have access to data related to popular research field with the aid of display technique. Therefore, how to quickly display the required information from the massive data is a challenge for the NSFC service system. We attempt to address this problem in the herein proposed cloud platform.

Usually, there are four ways to display resources in current platforms [6-11]. 1) Text display. As shown in Figure 2, displaying resource through text with illustrations and videos is the most common way, and it is simple, clear, comprehensive and direct. 2) List display. This display can further divided into information list and link list. The former s convenient for users to export the data which has been retrieved in several formats, CSV, Excel, XML etc., and make it easy to query and use in future. The latter is flexible and changeable, mainly provides users with relevant content links. 3) Resource Summary。 This way of resources display is also quite common, generally includes the name, the summary, the author, the date and such basic information, also some pictures when necessary. This way is generally in the form of lists, it usually has two forms that are no picture and picture with text. 4) Resource Presentation Based on Geographic Information. To query resource information according to geographic information, users can view the area of research and statistical information through click on a map of the area. This way of visual display is very expressive, and can meet the demand of a wide range of users.

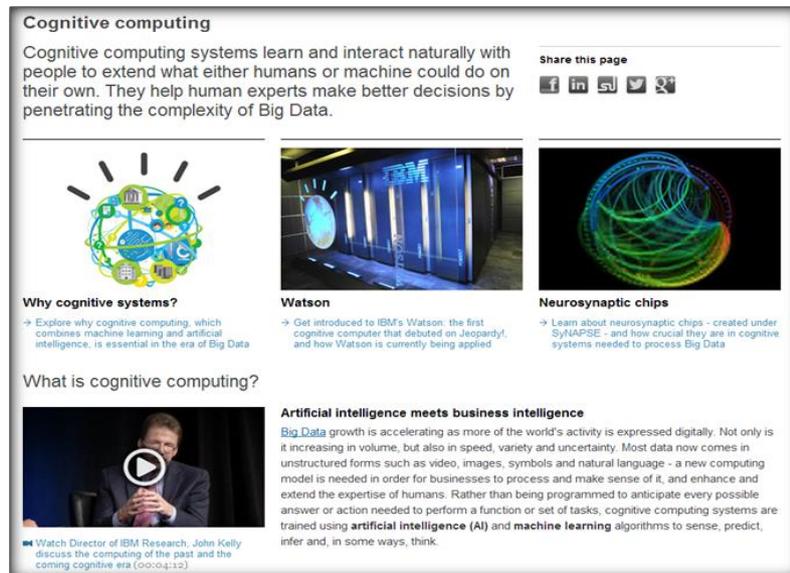


Figure 2. Illustration of Text Display

2.4. Function Service

In order to satisfy various demands from different users and the administrators, NSFC Service System has a variety of service, like information searching, data storage, data sharing, new and announcement releasing and display etc. To rapidly provide quality service, we implement the above-mentioned functions in our cloud platform through cloud computing and storage techniques. Thus, it is necessary to briefly introduce the main function service of our cloud system.

According to user needs of the science fund sharing service network, the primary service which is provided by website above can be summarized as follows[12-15]:

1) Information Resources Retrieval

Searching function is the first and basic function for information service website. As the data is becoming more and more huge, searching function is provided extensively by such website. How to learn the user needs in a simple, accurate and fast way and improve the searching results with a proper order is a crucial problem. Searching function can be divided into fast searching and advanced searching, aim to help users get information more conveniently. In addition, searching function can be implemented only through a search-box, which makes it easier. Users can further lock information by using second time searching when there are too many results after the first time searching.

2) Classified Resources Display

Information resources are displayed according to certain categories, including research areas, geographic information and the existing resources. Different searching classifications can be provided on different sections or pages, so that users can study them from multiple angles and dimensions. How to find and build the internal relations between existing resources and users' various needs is the key problem.

3) Statistical Display

Whether it is a search page or a browse page, the combination of statistical information and resources display can provide not only the static data, but also the dynamic and comprehensive data according to browsing and retrieving information from users, according to different basis, screening conditions and limitations. What's more, it can balance the relationship among the project, the funds, the results and the time well enough.

4) Relevant Policies

The Online grants management for NSF (research.gov) provides a policy electronic library for the public, scientific research organizations and policy makers to realize full text retrieval.

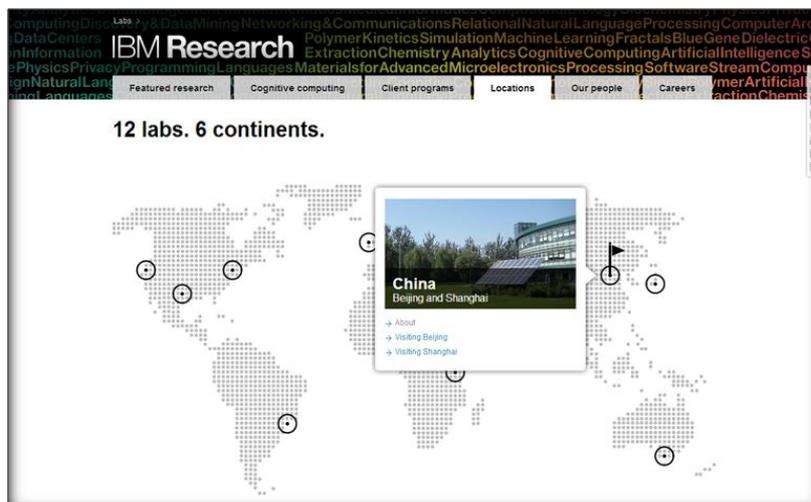


Figure 3. Display based on Geographic Information

3. Cloud Platform Architecture

Infrastructure platform includes three parts of equipment environment, private cloud and system software, as shown in Figure 4. Among them, the private cloud is the core of the infrastructure platform, which uniformly manages and assigns tasks to the computing components, storage components and network components, as well as provides necessary underlying interface support for the system software.

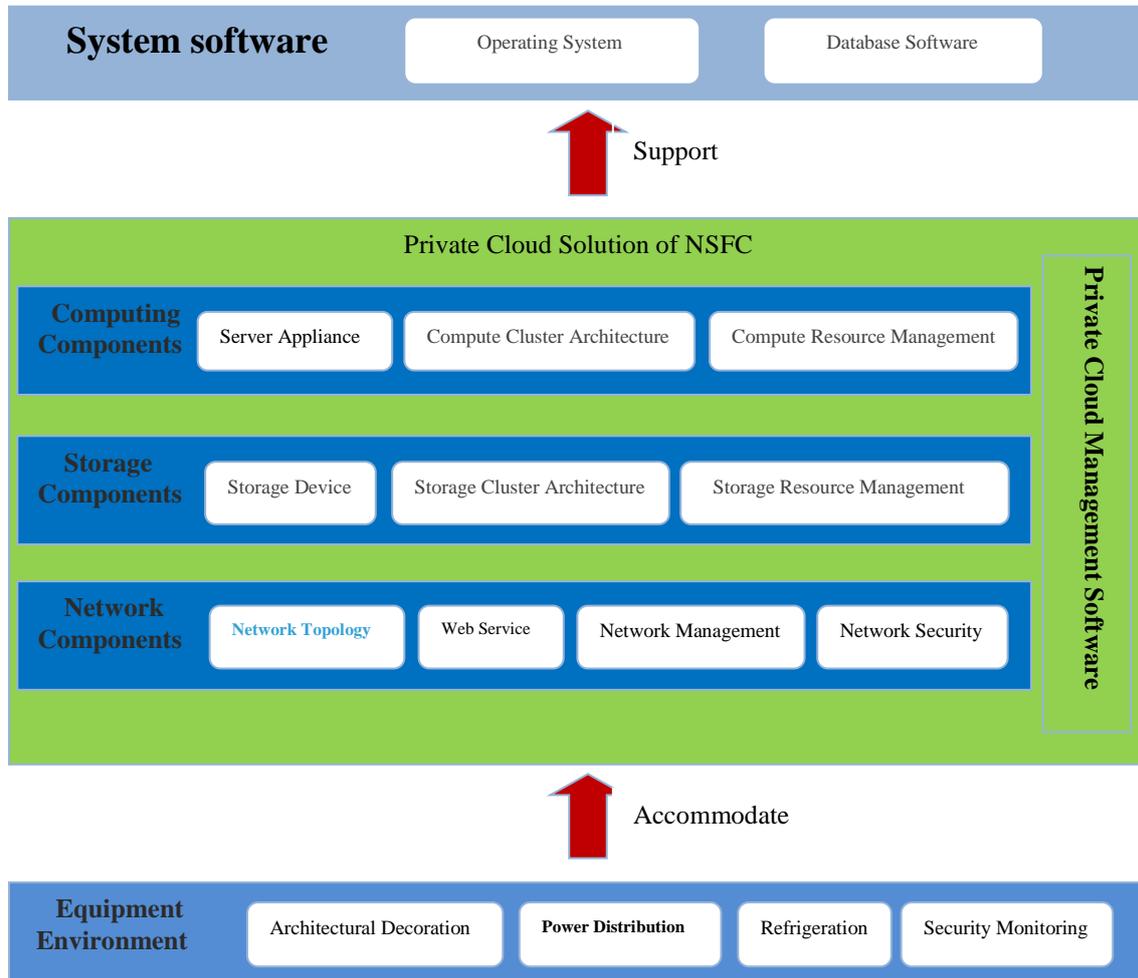


Figure 4. Cloud Platform Architecture

4. System Software and Equipment Environment

4.1. System Software

System Software mainly contains two parts: Operating System (OS) and Database Management System (DMS). At present, the mainstream of OS includes UNIX, Linux, Microsoft Windows, VMware, etc. And DMS mainly includes Oracle DB, Microsoft SQL Server, IBM DB2, etc. In this project, we use UNIX and Linux as the OS for DB running, because of their characteristics of security, stability, high-performance, maintainability and low-cost. And Oracle DMS is suggested.

OS is a support system in which basic scientific research knowledge service system can perform some applications such as data sharing and service. Special management personnel should maintain this basic scientific research knowledge service system, and update the system periodically, fix the vulnerabilities, monitor and optimize the performance of OS, DMS and other application software. Our management structure of

infrastructure platform DB is to place centralization first, distribution second. Meanwhile we strictly follow international open standards and criterion. Infrastructure platform should perform essential monitoring and optimization to DB. By this the retrieve speed and strategy of backup recovery will be improved.

4.2. Equipment Environment

To construct more safe and reliable equipment environment, the following important factors should be taken into consideration: decoration materials, suspended ceiling, floor, walls, air conditioner, exhaust system, illumination system, doors and windows, firefighting system, grounding scheme, other equipment and auxiliary materials, CCTV image monitors, environment monitors and so on. The concrete contents of the construction are shown in Table 1. And Table 2 shows the functional requirements of equipment environment construction.

Table 1. The Contents of Equipment Environment Construction

Contents	Description
Equipment Environment	Building decoration
	Power supply system
	Constant temperature system
	Fire protection system
	Lightning protection system
	Ventilation system
	Monitoring system
	Access control system
	Generic cabling system
	System integration

Now NSFC has completed the construction of data center equipment system. To meet the needs of our project, based on the existing environment conditions, the construction of equipment environment requires us to do following: appraising existing environment according to PRC national standard “Design Specifications of Electronic Information Systems Room GB50174-2008”, reforming and perfecting the central equipment room in accordance with this national standard.

Table 2. The Contents of Equipment Environment Construction

module	requirements	function description
Equipment environment	Conform to the standard specification	Equipment environment should be constructed according to various standards;
	Reasonable building environment	Space resource must be utilized adequately; and weight capability of building must conform to requirement;
	Excellent equipment environment	Maintain a constant temperature and humidity, anti-static and dustproof storage environment in quite long time;
	Redundant electric power	Redundancy backup scheme of power supply should be adopted;
	Reliable security system	Keep 24-hr monitoring, 7 days a week, and preclude all irrelevant personnel by using access control system.

5. Network Components

The planning of network resources needs to not only emphasize the non-blocking design of network data plane, but also ensure the three layers of protection system of the network, namely, “the big two layers”. Each internal host ensures minimum grant access through the host firewall to constitute the inner protection, using the business system as a unit, in the business system. Each business system is treated as an independent VLAN to build the second protection among the business systems. The system only opens the minimum service port according to the principle of minimum authorization, and ensures the security of network boundary via the outer firewalls to constitute the most outer layer protective system. The three layers based network protection system should be able to prevent the attacks of network level.

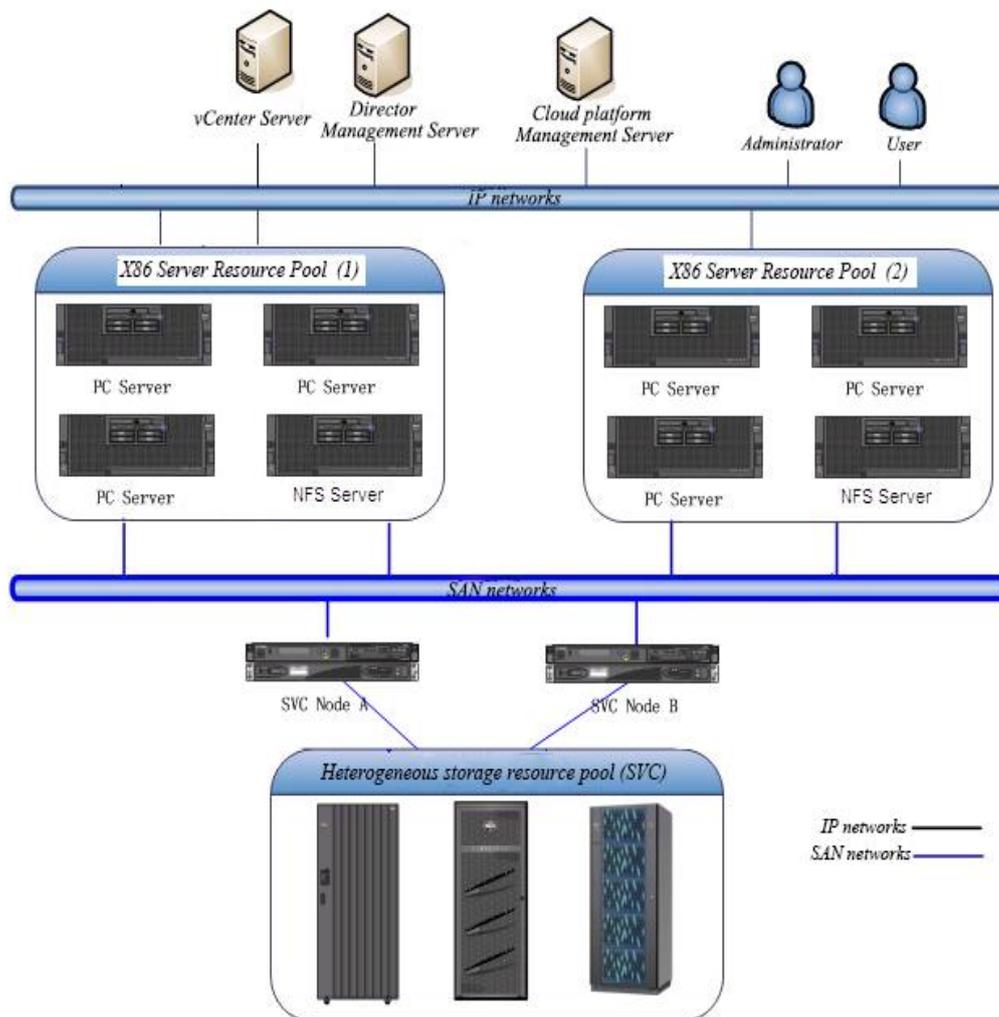


Figure 5. Physical Topology of Networking on Infrastructure Platform

It is better to adopt the platform with unified application delivery, if the security protection on application layer is needed. This platform is analogous to load balancing. The external service of each business system is registered on the unified application delivery platform, including domain name which is bound to public IP of a unified application delivery platform, IP address mapped to the Intranet and Intranet server port. All the requests from users for the service system can be forwarded by unified application delivery platform. Therefore, the users can't reach the business system directly on application layer, which will reduce the risk of

immediate penetration. In addition, the requests of users and business response are fully recorded by the unified delivery platform. It not only can find and alarm the potential intruders on application layer through the behavior analysis, but also can block them in time by configuring the blacklist. The Principles of networking on infrastructure platform are shown in Figure 4.

The entire cloud computing platform, including the network interface, network link and the key network equipment, needs to configure the redundant components. On the network interface, each physical server needs to configure at least three network cards respectively for the business services, virtualization platform hosting management and IP storage system interconnection. The business services network is divided into the public network area, Internet access area and private network area by the MPLS VPN, According to the difference of business properties. Virtualized computing resources enable to migrate freely in different network areas.

6. Private Cloud Construction and Storage Components

6.1. Private Cloud

Wherever Times New Roman is specified, Times New Roman may be used. If not available in your word processor, please use a font closest to Times New Roman that you have access to. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

Table 3. Construction Contents

Construction contents	Task description
Network components	The architecture design of Network facilities
	Set up of Network facilities
	The network monitoring and the management system
	Firewall
	The architecture design of Storage facilities
	Database cluster
	Management system of database cluster
	Distributed file system
	Backup Disaster Recovery Systems
Computing component	The architecture design of computing facilities
	Set up of computing facilities
	Calculation of Resource Management System
	Virtualization cluster
Private cloud	Construction of private cloud hardware
	Construction of private cloud management software

The construction of the NSFC internal cloud means building the software and hardware architecture. The architecture provides finite people in the internal firewall with hosting services. Private cloud made accessible to the specific clients for the control of data, security and the quality of service effectively. There are many solutions including the open source and the commercialize of private cloud can be chosen. The commercializing of private cloud is adopted owing to its mature and reliable after considering the subsequent demands of the systems and services. The construction of the internal cloud considers the building task of the software and the hardware, shown in Table 3.

6.2. Storage Component

To share the storage platforms is the key project of constructing private cloud. Once the problem occurs to the storage, it will directly influence the online business of private cloud and bring the negative effect for a long period. The cause of the problem is the use of centralized storage mode. There are two questions under consideration when planning the storage resources. How to satisfied the requirements of the storage capacity and performance on the business. How to implement the function of data backup and disaster recovery? The design scheme of disaster recovery architecture and system operation model should be proposed according to the characteristics of this project.

7. Computation Components

Because the computing resource are related in the selection of server, so the mainframes and the minicomputers need not be considered in this project, instead suggest considering eight rack servers, four rack servers, two rack servers, blade servers etc..In practice, the choice between the blade server and the rack server can consider the factors such as space, refrigeration, power, cost, use of the scene and so on. Usually, if you choose blade server that it will save space, but the power consumption per unit space of it is relatively large. Compared with the former, the power consumption of the rack server can be controlled , but it will occupy more space. The design of computing resource pool needs to start from two aspects: the cluster architecture and the server configuration.

7.1 Cluster architecture

According to the purpose, the server can be divided into a plurality of computing cluster units.

- (1) Virtual cluster unit: The virtual cluster unit is used to support the hosting business. It tries to use N+1 or N+2 model to build the cluster as far as possible, that is, N production servers, 1 to 2 sets of alternate migration server.(The choice of N is determined by the performance of spare nodes and the failure rate of the production nodes).
- (2) Database cluster unit: The database cluster unit is used to support massive data processing environment, the choice of the cluster can be determined by the database platform. For example, for MySQL and SQL Server, you can build a "single-master and multi-slave" cluster; for Oracle, you can build a RAC cluster.
- (3) Computing cluster unit: computing cluster unit is the smallest unit in the computing pool, and a computing pool is composed of a plurality of computing cluster units which are used for different purposes.

In addition, according to its performance, the computing cluster unit can be divided into high performance computing environment, medium performance computing environment and low performance computing environment , and it can meet the needs of different SLA.

7.2 Selection of the Server Configuration

For the server configuration that supports virtualization, the following issues should be considered:

- (1) CPU capacity: The selected chip can support virtualization.

- (2) Memory capacity: The ratio of memory and CPU usually is 4:1 , that is, one CPU kernel correspond 4GB memory, you can also choose a higher ratio such as 6:1 or 8:1.
- (3) Local disk: If you rely mainly on shared disk array, the local disk does not need too much, you can choose two disk, and just do RAID1 as long as the installation of virtual platform can be completed.
- (4) Network card: The network card should be configured at least four Gigabits: one for the platform management, one for the virtual machine migration, two for the virtual machine business.
- (5) HBA card: HBA card can be used to connect the FC storage networks, it generally needs to configure the dual HBA interface to connect the high performance storage array.

8. Conclusions

The Service System of NSFC Knowledge Base is a platform for the National Natural Science Foundation of China, serving the majority of scientific researchers, institutions and NSFC itself. As rapidly increasing of NSFC foundations, NSFC has accumulated massive academic and management data on the projects, experts, institutions and achievements and varieties of project reports. In order to better manage the NSFC project-relevant digital data and offer a friend platform and high-quality server, we develop a new cloud platform for the NSFC basic information system, which mainly includes the parts: computer room environment, private cloud, system software, storage components and computing components. Future work will focus on improving the security of this platform.

Acknowledgments

This paper is a revised and expanded version of a paper entitled "Construction Scheme for Cloud Platform of NSFC Information System" presented at ISI 2016, August 19-20, Harbin, China.

References

- [1] B. Y. Wang, M. Li and H. T. Wang, "Geometric range search on encrypted spatial data", IEEE Transactions on Information Forensics and Security, vol. 11, no. 4, (2016), pp. 704-719.
- [2] K. Djemame, D. Armstrong and J. Guitart, "A risk assessment framework of cloud computing", IEEE Transactions on Services Computing, vol. 4, no. 3, (2016), pp. 265-278.
- [3] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun and K. Ren, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing", IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, (2016), pp. 2594-2608.
- [4] <http://www.nsf.gov.cn/publish/portal1/tab157/> May, (2015).
- [5] <http://npd.nsf.gov.cn/> May, (2015).
- [6] <http://www.doaj.org/> Retrieved 1 May, (2015).
- [7] <http://www.highwire.org/lists/freeart.dtl> Retrieved 1 May, (2015).
- [8] http://en.wikipedia.org/wiki/Open_access Retrieved 1 May, (2015).
- [9] "Read the Budapest Open Access Initiative", Budapest Open Access Initiative, Retrieved 1 May, (2015).
- [10] K. G. Jeffery, "Open Access: An Introduction", Ercim News, (2006).
- [11] <http://www.nsf.gov.cn/publish/portal1/> Retrieved 1 May, (2015).
- [12] www.research.ibm.com Retrieved December, (2015).
- [13] <http://www.nsf.gov.cn/publish/portal1/tab285/> Retrieved December, (2015).
- [14] <http://or.nsf.gov.cn/> Retrieved December, (2015).
- [15] J. Li, M. Zhang, D. Li, W. Zhang and J. Wang, "Construction Scheme of NSFC Open Access Library", International Journal of Security and Its Applications, vol. 9, no. 8, (2015), pp. 243-252.

Authors



Jianjun Li. He received the B.S. degree from Beijing Forestry University in 1989. Now, he is the director of the Information Center of National Natural Science Foundation of China (NSFC). His main research interest includes NSFC information system management ranging from system architecture design, construction, operation, database maintenance, and security etc.



Yuhui Zheng. He received the B.S. and Ph.D degree from Nanjing university of Science and Technology. Now, he is associate professor at the School of Computer and Software, Nanjing University of Information Science and Technology. His research interest covers image processing, pattern recognition, and remote sensing image restoration. He is a member of the IEEE.



Jin Wang. He received the B.S. and M.S. degree from Nanjing University of Posts and Telecommunications, China in 2002 and 2005, respectively. He received Ph.D. degree from Kyung Hee University Korea in 2010. Now, he is a professor at the School of Computer and Software, Nanjing University of Information Science and Technology. His research interests mainly include routing protocol and algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a member of the IEEE and ACM.