Astroinformatics
Proceedings IAU Symposium No. 325, 2016
M. Brescia, S.G. Djorgovski, E. Feigelson,
G. Longo & S. Cavuoti, eds.

doi:10.1017/S1743921316012709

Design and Implement of Astronomical Cloud Computing Environment In China-VO

Changhua Li[1], Chenzhou Cui[1],Linying Mi[1],Boliang He[1], Dongwei Fan[1],Shanshan Li[1], Sisi Yang[1], Yunfei Xu[1], Jun Han[1], Junyi Chen[2], Hailong Zhang[3], Ce Yu[4], Jian Xiao[4], Chuanjun Wang[2], Zihuang Cao[1], Yufeng Fan[2], Liang Liu[6], Xiao Chen[7], Wenming Song[7] and Kangyu Du[8]

[1]National Astronomical Observatories, Chinese Academy of Sciences (CAS), 20A Datun Road, Beijing 100012, China
email: lich@nao.cas.cn

[2]Yunnan Astronomical Observatory, CAS, P.0.Box110, Kunming 650011, China
email: wcj@ynao.ac.cn

[3]Xinjiang Astronomical Observatory, CAS, 150 Science 1-Street, Urumqi, Xinjiang 830011, China
email: zhanghailong@xao.ac.cn

[4]Tianjin University, 92 Weijin Road, Tianjin 300072, China
email: xiaojian@tju.edu.cn

[5]Purple Mountain Observatory, 2 West Beijing Road, Nanjing 210008, China
email: liangliu@pmo.ac.cn

[6]Shanghai Astronomical Observatory, 80 Nandan Road, Shanghai 200030, China
email: cx@shao.ac.cn

[7]Beijing University Of Technology, Beijing 100124, China
email: songwenming@nao.cas.cn

[8]Central China Normal University, 152 Luoyu Road, Wuhan 430079, China
email: dukangyu@nao.cas.cn

Abstract. Astronomy cloud computing environment is a cyber-Infrastructure for Astronomy Research initiated by Chinese Virtual Observatory (China-VO) under funding support from NDRC (National Development and Reform commission) and CAS (Chinese Academy of Sciences). Based on virtualization technology, astronomy cloud computing environment was designed and implemented by China-VO team. It consists of five distributed nodes across the mainland of China. Astronomer can get compuitng and storage resource in this cloud computing environment. Through this environments, astronomer can easily search and analyze astronomical data collected by different telescopes and data centers , and avoid the large scale dataset transportation.

Keywords. cloud computing, virtual machine, astronomy, migration computing, cloudstack

1. Introduction

Recent years, with the coming of many large telescope, the daily increment of data increase to TB level. Furthermore, with the good progress of many international cooperation astronomy project, such as TMT, LSST, SKA, etc. The daily increment of data will increase to PB level in the near future. Astronomy research has step into the big data era. The large scale data put forward great challenges on processing, transportation, analysis of data. With the limitation of network bandwidth and computing, storage ability of private computing of astronomer, the traditional research mode would be disability.

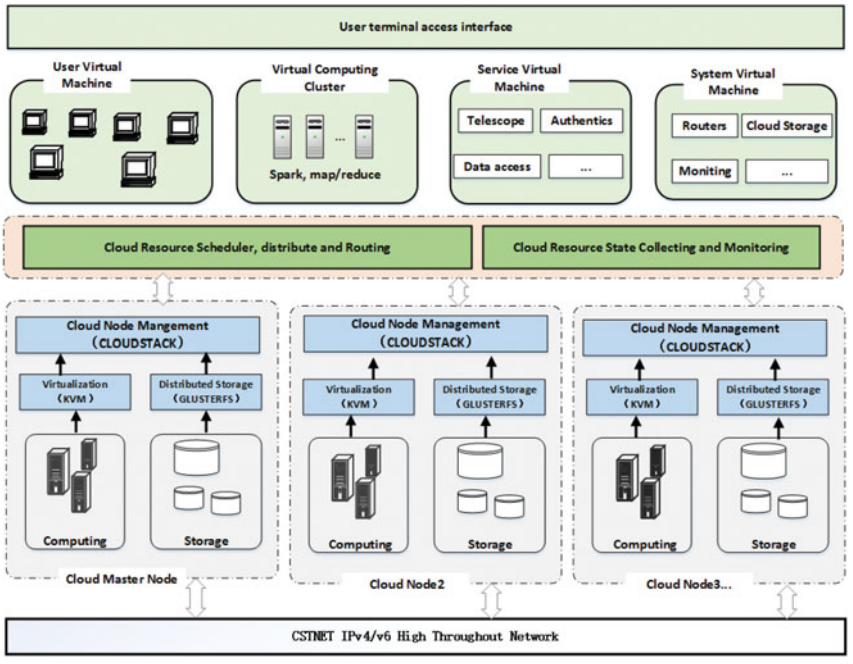

Figure 1. The framework of astronomy cloud computing environment

Migration computing will become this new solution. The construction and application of cloud computing environment would make this new solution reality. Astronomy cloud computing environment would become a new e-science platform. Cloud computing is to integrate geographical distribution computing resource through network and to provide computing and service on demand based on virtualization technology. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction(Mell,P. & Grance,T. (2011)). According to the provisioned resource type, Cloud Computing can be divided to three different type: IAAS means Infrastructure as a service, to provide hardware service on demand, PAAS means Platform as a service, to provide application software service over the internet, SAAS means software as a service, to provide browser access of software. Virtualization is the key technology of cloud computing that combines or divides computing resources to present one or many operating environments using methodologies like hardware and software partitioning or aggregation, partial or complete machine simulation, emulation, time-sharing, and others(Susanta Nanda & Tzi-cker Chiueh).

2. The Architecture of Astronomy Cloud Computing Environment

The telescope is the center and basis of astronomy research, and the main source of data. The big data from telescope was saved in the nearest data center, which become a basic unit of cloud computing platform system, was called cloud computing node. Each node have enough computing and storage resource, and is also a complete cloud computing environment. Figure 1 shows the framework of astronomy cloud computing environment.

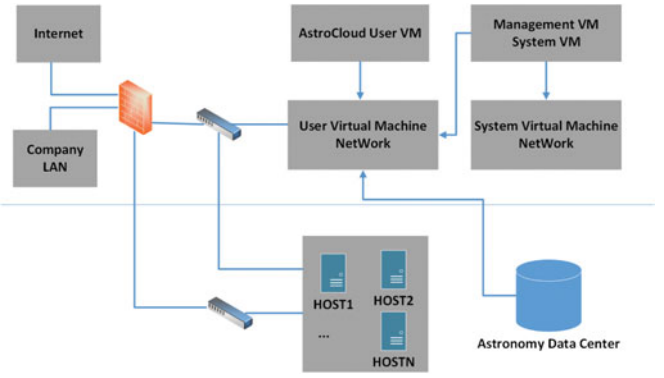

Figure 2. The topology of networks in a single astronomy cloud node

Astronomy cloud computing environment uses the passport of CSTNET as the third certification system. Astronomer can directly login cloud environment through their email account of CSTNET. The overall system include five cloud nodes, which are the data center of NAOC, SHAO, PMO, XAO, YNAO, and integrated the astronomy large scale data storage system, which provide global data storage service and private file storage. In addition, in the region resource schedule layer, we provide many system management components, such as the user dashboard, usage statistics, template synchronization, etc. CloudStack is an open source software platform that pools computing resources to build public, private, and hybrid Infrastructure as a Service (IaaS) clouds. CloudStack manages the network, storage, and compute hosts that make up a cloud infrastructure. We can use CloudStack to deploy, manage, and configure cloud computing environment. Cloudstack is the core component of astronomy cloud computing environment. It manages all hosts, storage, and network of each node.

3. Design of astronomy cloud environment network

Internal network is an important part of cloud, which is much more complicated than the official LAN. According to the function of network traffic, CloudStack divides 4 network traffic type, namely public, management, storage, guest and two network design model: basic network and advanced network. Basic network model can only have one guest network, all traffic type share a big address space. Advanced network model allows for the creation of multiple guest networks with a tagged VLAN. Under astronomy cloud environment, we created a shared guest network with a tagged VLAN 1 as the default network of virtual machines. We can also create more isolated guest network to meet the security requirement for some special projects. Figure 2 shows the connectivity of physical and logical resource. Some data resource and service, out of the astronomy cloud environment, can also connect to the shared guest network. It is very convenient to access the big astronomy dataset for guest virtual machine.

4. The access way of virtual machine

Because the virtual machine is in a virtual environment, astronomer cant see it, cant manage and program debug like in a physical server before. A convenient, high efficiently access way to virtual machine was vital to promotion of cloud environment. In astronomy cloud computing environment, we integrated the VNC service and webVNC plugin to redirect the standard output of virtual machine to browser. Astronomer can operate his

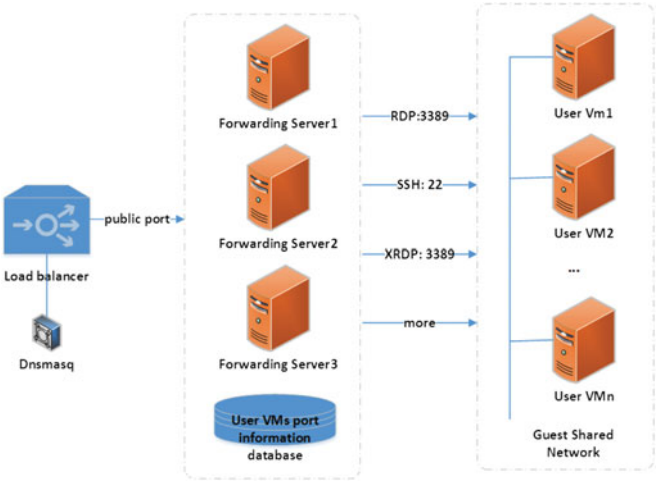

Figure 3. The topology of networks in a single astronomy cloud node

virtual machine by internet browser. But this method has two disadvantages. One is the display effect depends on the internet speed; Aother is the disability of copy and paste operation, which is very inconvenient. So, we integrated the NAT and port forwarding function of iptables to provide port forwarding service. Astronomer can login virtual machine by client of SSH, RDP. SSH and RDP are the default protocol, when virtual machine create, the port forwarding setting could be configured automatically. Figure 3 shows the workflow of access virtual machine. In order to avoid the single point faults of forwarding server, based on the Dnsmasq, a lightweight domain name resolution service, we implements the load balance of forwarding service. The relation of public port and private port of virtual machine saved in a database.

5. Conclusion

In general, astronomy cloud computing environment promotes the progress of astronomy e-science under the big data era, ease the burden of IT management on astronomer. Meantime, cloud computing environment improves resource utilization rate, reduce the cost of IT infrastructure. In the future, we will integrate the Hadoop, Spark computing framework in the astronomy cloud environment to enable the large scale data processing service.

Acknowledgements

This paper is funded by National Natural Science Foundation of China (11503051, U1231108, U1531246, U1531115), Ministry of Science and Technology of China (2012FY120500), Chinese Academy of Sciences (XXH12503-05-05). Data resources are supported by Chinese Astronomical Data Center.

References

Mell, P. & Grance, T. 2011, *NIST Special Publication*, 800, 145
Susanta Nanda, Tzi-cker Chiueh *http://www.ecsl.cs.sunysb.edu/tr/TR179.pdf*