Docker based Datacenter for Grid Computing in GSDC HTC Environment

Jeong-Heon Kim, Heejun Yoon, and Duseok Jin

Abstract—We introduce Docker based datacenter for grid computing, which is used by Global Science Experimental Data Hub Center of Korea Science and Technology Information Institute. Our computing groups fall into three categories depending on their usage characteristics. The categories are physical group, static group, and dynamic group. Physical group are fixed and difficult to change. Nevertheless, some special applications must run on the physical group. Static group and dynamic group are implemented using Docker container virtualization. Static group are beneficial for less changing usage characteristics that are common to most of the resources for grid computing. We maintain simple and flexible management for most applications with a static group. Dynamic group support lively changes while operating with the implementation of additional scheduler and executor.

Research Keywords—Container, Datacenter, Grid Computing, Management

1 INTRODUCTION

The Global Science Experimental Data Hub Center (GSDC) of Korea Institute of Science and Technology Information (KISTI) supports various researches through applications using grid computing and cluster computing resources [1]. We often change the organization of computing resources to support various researches. Re-organization of computing resources can result in complex management tasks. The operating automation is proposed as a solution to management complexity [2], [3], [4].

We can focus on software distribution using the Linux container cloud without reinstalling the operating system. This allows us to maintain simpler operating automation in the Linux container cloud. However, special applications require a physical machine or a specific Linux kernel. In addition, the usage characteristics of the elements that make up the grid computing are different from the general usage characteristics of the cloud environment. This paper presents the organization and structure of computing resources based on Linux container cloud for GSDC-KISTI datacenter.

- *Jeong-Heon Kim is with the Department of National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information, Daejeon, Korea 34141. E-mail: jh.kim@kisti.re.kr.*
- *Heejun Yoon is with the Department of National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information, Daejeon, Korea 34141. E-mail: k2@kisti.re.kr.*
- *Duseok Jin (corresponding author) is with Department of National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information, Daejeon, Korea 34141. E-mail: dsjin@kisti.re.kr.*

2 CHARACTERISTICS OF NODES IN GSDC

In order to function as a grid member for international cooperation, we need close cooperation with a central resource broker. As shown in Figure 1, each component of grid computing should be routed directly to external components. As represented in Table 1, 95% of the elements should be registered in the public do-

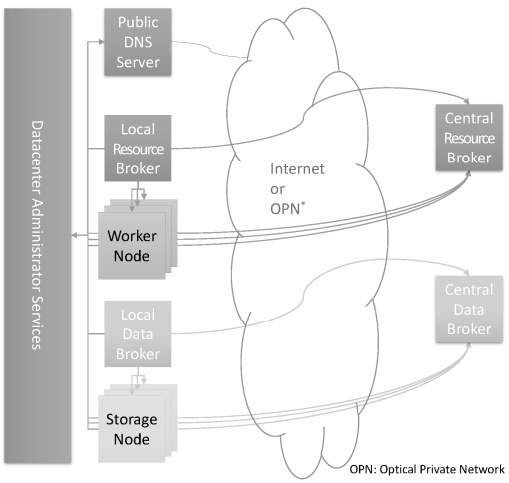

Fig. 1. Datacenter components for grid computing.

Table 1. Properties of Datacenter Components for Grid Computing

Type	**Broker**	**Worker**	**Storage**	**Admin**
Share	5%	70%	20%	5%
Host Utilization	Low	Full	Full	Mid
Certification Requirement	Required	Required	Required	Required
Public DNS registration	Required	Required	Required	Not need
Public IP Address Requirement	Required	Required	Required	Not need

main name service and have a public Internet address. In addition, since 90% of the components use most of the host resources, 1 physical node - 1 virtual node - 1 public address is formed.

3 Organization of Elastic Datacenter for Grid

Depending on the characteristics of resources for grid computing, most of the resources belong to static groups. Even a static group with little change can be adjusted once per quarter due to equipment changes and once a month due to hardware failure. Figure 2 shows the organization of elements for managing with each computing group. Therefore, administrations are divided into two groups for managing physical resources and managing containers. Figures 3, 4, and 5 show two separate container clusters and services separated for two groups of administration services.

4 Conclusions

The components for grid computing show different characteristics from those for general web services. In particular, we can see the difference in network configuration and resource deployment. Therefore, we need specialized network configuration and resource organization. Based on the operational data of GSDC, we made a suitable organizational structure. In the future, we would update the organization with operational data based on the proposed organization.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) through contract N-17-NM-CR01 and the Program of Construction and Operation for Large-scale Science Data Center (K-17-L01-C05).

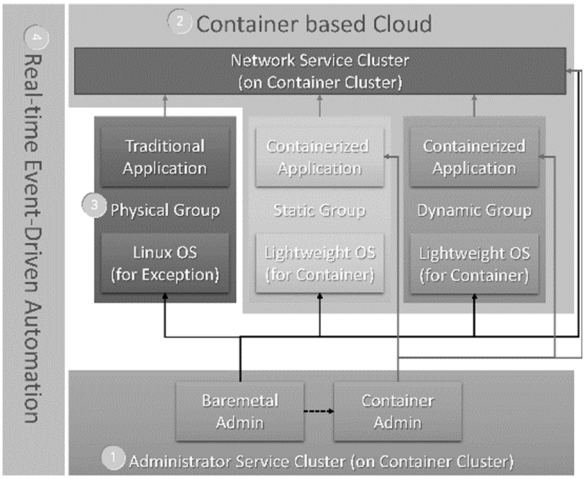

Fig. 2. Organization of container based cloud for grid.

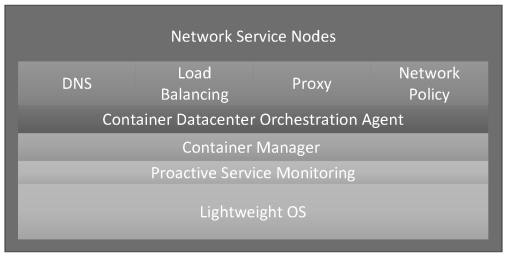

Fig. 3. Network services on a datacenter service container cluster orchestration.

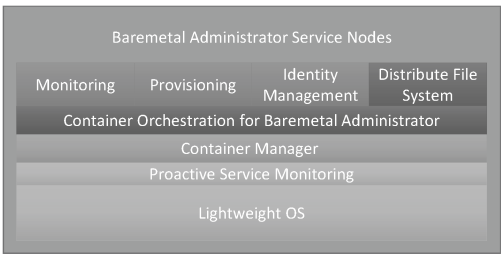

Fig. 4. Baremetal administration services on an initial container cluster.

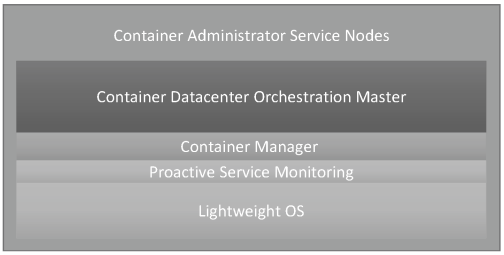

Fig. 5. Structure of datacenter services container administration node.

REFERENCES

[1] S.U. Ahn, I.Y. Yeo, and S.O. Park, "Secure and Efficient High-Performance PROOF-Based Cluster System for High-Energy Physics," J. Supercomputing, vol. 70, no. 1, pp. 166-176, Oct. 2014.

[2] H. Yoon, I.Y. Yeo, J.H. Kim, "Updating the Trusted connection of Re-Organized Computing Resource Under the Automated System Management Platform," J Supercomputing, vol. 70, no. 1, pp. 200-210, Oct. 2014

[3] C.W. Kim, H. Yoon, D. Jin, S.O. Park, "Integrated Management System for a Large Computing Resources in a Scientific Data Center," J Supercomputing, vol. 72, no. 9, pp. 3511-3521, Sep. 2016.

[4] J.H. Kim, D. Jin, P. Lee, "Cyber Physical System-Based Convergence Operation of Data Intensive Computing Resources," Wireless Pers. Commun., vol. 89, no. 3, pp. 881-891, Aug. 2016.

Jeong-Heon Kim received his B.Sc., Engineer diploma from Chung Ang University (Seoul, Korea) in 2003, and his Ph.D. in computer science and engineering from Chung Ang University in 2013. Since 2013, he has been worked as a researcher in the Korea Institute of Science and Technology Information. He is currently interested in Image Processing and System management.

Heejun Yoon received the M.S from the Computer Engineering at Chungnam National University, KR in 1997. He has been serving as a Senior Researcher at Korea Institute of Science and Technology Information since 2000. His research interests include Large data processing, Parallel Computing, Cyber-Infra system, Computer Education.

Duseok Jin is a principal researcher at National Institute of Supercomputing & Networking, KISTI. He received his M.S in Computer Science from Univ. of Chonbuk, Korea and his Ph. D. in Computer Science from Univ. of Paichai, Korea in 2001 and 2011 respectively. His Research interests are Information Retrieval system, Cloud Storage and parallel/distributed file system.