International Journal of u- and e- Service, Science and Technology
Vol.9, No. 3 (2016), pp.9-16
http://dx.doi.org/10.14257/ijunesst.2016.9.3.02

Developing a Decision Support Model for a High Performance Computing System Selection for Computational Science using AHP

Namgyu Kim[1], Hyoung Woo Park[2], Sudong Park[3] and Hyungjin Lee[4,*]
[1,2,4]Korea Institute of Science and Technology Information, South Korea
[3]Institute for Basic Science
[1]ssgyu@kisti.re.kr, [2]hwpark@kisti.re.kr, [3]triznik@ibs.re.kr, [4]longleg@kisti.re.kr

Abstract

Traditionally computational scientists have used supercomputers to solve their scientific problems using multi-processors and large-scale shared memory, and scientific communities have also used Grid computing for scientific projects using large-scale computation and data resources. Recently Cloud computing is a emerging infrastructure to be considered for scientific applications and several institutes applied this to their projects. But there is no sure method for computational scientists to select a environment to fit their researches though diverse researches using these infrastructures are conducted in a variety of science domains. In this paper we describe three infrastructures for scientific researches through literature reviews, and we deliberate on what factors are considered to select one infrastructure to solve scientific problems from a scientists' point of view. And then, analytic hierarchy process is introduced as an approach for choosing a proper infrastructure, and we propose a model to choose a appropriate infrastructure reflected by various aspects of computational science's properties via this approach. Thus this paper offers a novel viewpoint to scientists when they choose a high performance computing environment for their researches.

Keywords: *HPC, Computational Science, AHP*

1. Introduction

Traditionally supercomputers using multi-processors and large-scale shared memory are representative facilities for computational science which is using computers to analyze and solve scientific problems with mathematical models. Also Grid computing infrastructures using geographically dispersed and large-scale computation and data resources such as Teragrid, LHC Computing Grid, Open Science Grid and so forth have been operated for large-scale collaborative scientific projects. Recently Cloud computing has be highlighted in commercial areas as a emerging technology and several scientific projects have tried to apply this to scientific applications and a case in point is NASA.

But although various researches using these high-performance computing infrastructures are conducted in a variety of science domains, when computational scientists select a proper infrastructure to fit for their researches it is hard to find selection methods with appropriate criteria except for a performance and scale of computing resources and storage volumes.

In this paper we examine three infrastructures for scientific researches through literature reviews, and then we take into account what factors should be considered from a scientists' point of view when scientists select an infrastructure to solve their scientific problems. And we introduce Analytic Hierarchy Process(AHP) as an approach for choosing a proper infrastructure, and we propose a model of a appropriate infrastructure

---

[*] Corresponding Author

ISSN: 2005-4246 IJUNESST

selection reflected by various aspects of computational science's properties via AHP. Our key contribution is to offer a novel viewpoint to scientists when they choose a high performance computing environment for their researches.

The remainder of the paper is consists of following sections. In related work section three infrastructures we mentioned above are identified and described briefly, and then compared with several aspects. In AHP section AHP methodology is introduced and described. And then in Model section we propose our model to select a infrastructure, and in Case Study section we show a simple example applied with the proposed model. Finally we summarize our study and report research limitations, and then describe future works.

2. Related Works

2.1. Supercomputer

Generally a supercomputer is defined as a high-end computer at the frontline of current processing capacity, by speed of calculation and storage capacity. At present the systems in the Top 500 supercomputer list released twice a year by Top500.org are designated, and they are considered as the most powerful general purpose computing systems that are in common use for high end applications around the world.

Commonly the types of supercomputers are Vector supercomputers, Symmetric multiprocessor systems(SMP), Massively parallel processing systems(MPP), and Cluster systems. But at present only MPP and Cluster systems are in the Top 500 supercomputer list, thus only MPP system and Cluster system are described briefly. MPP system is a architecture that each CPU has each dedicated memory and its own operating system to handle a program, and many CPUs work in parallel to execute the same program. Thus for collaboratively processing the program by CPUs a messing interface like Message Passing Interface(MPI) must be necessary to handle CPUs between the different CPUs belong to the MPP system.3 Cluster system is a collection of computers that are interconnected within a private local area network such as gigabit Ethernet, SCI, Myrinet and Infiniband. Each compute node(computer) may have different system specifications and access to various types of storage devices within a dedicated network made up of high-speed and low-latency system of switches. This approach is to design an efficient computing platform that uses a group of commodity computer resources to improve the performance and availability orchestrated by clustering middleware, a software layer to form a single computer resource.

2.2. Grid Computing

Grid Computing is aim to enable resource sharing and coordinated problem solving to coordinates resources by decentralized control, using standard, open, general-purpose protocols and interfaces in dynamic, multi-institutional virtual organizations. The concept of Grid computing is based on using the Internet as the wide spread availability of powerful computing resources with low-cost commodity computers to deliver qualities of service according to their availability, capability, performance, and users quality-of-service requirements. For this Grid computing has a middleware to divide and assign fractions of a job among resources using grids account for different administrative domains with access policies. But this guarantees the secure access by user identification and the users and the institutions define their own rules for resource sharing in virtual organization.

Table 1. Compasion of Supercomputer, Grid and Cloud Computing

Attribute	Supercomputer	Grid	Cloud
Capacity	Fixed	Average to high; growth by aggregating indepently managed resources	High; growth by elasticity of commonly managed resources
Capability	Very high	Average to high	Low to average
Virtual Machine Support	Rarely	Sometimes	Alaways
Resource sharing	Limited	High	Limited
Built-in Workload Management	Yes	Yes	No
Distributed Workload Across Resources from Multiple Admin Domains	No	Yes	No
Interoperability	No	Average	Low
Security	Medium	High	Low
Allocation	Centralized	Decentralized	Both
Resource Handling	Centralized	Distributed	Both
Coupling	Tight	Loose/Tight	Loose
Protocols/API	MPI, Parllel Virtual	MIP, MPICH-G, GIS, GRAM	TCP/IP, SOAP, REST, AJAX
Reliability	No	Half	Full
User friendliness	No	Half	Yes
Virtualization	Half	Half	Yes
Standardized	Yes	Yes	No
Business Model	No	No	Yes
Task Size	Single large	Single large	Small & medium
SOA and heterogeneity support	Not Supported	Supported	Supported
Multitenancy	No	Yes	Yes
System Performance	Improves	Improves	Improves
Self service	No	Yes	Yes
Computaion service	Computing	Maximum Computing	On demand
Resource support	Homogeneous and heterogeneous (GPU)	Heterogeneous	Heterogeneous
Data Locality Exploited	No	No	Yes
Application	HPC, HTC	HPC, HTC, Batch	SME interactive appcliation
Switching cost	Low	Low	High
Value Added Services	No	Half	Yes
Network Type	Private, LAN	Private, WAN	Public, WAN
Resource reservation	Pre-resereved	Pre-reserved	On-demand
SLA constraint	Strict	High	High
User Interface	Single system image	Diverse and dynamic	Singel system image
Initial infrastructure cost	Very high	High	Low
Administrative domain	Single	Multi	Both

2.3. Cloud Computing

Cloud computing is a emerging model for Information Technology (IT) services based on the Internet to share resource pool of abstracted, virtualized, dynamically-scalable, managed, configurable computing resources such as storage, server, applications, and services with the ease-of-access to remote computing sites using the Internet. And this model provides the distinguished feature to provision rapidly and release resources to be delivered on demand to external customers via the Internet with minimal management effort by service provider. And also Clouds often offer users a single point access, the QoS requirements of customers, and typically including SLAs for computing the consumer needs.

Generally the types of Cloud computing is classified depending on the types of serviced resources and three kind of service are separated principally: Software as s Service(SaaS), Platform as a Service(PaaS), and Infrastructure as a Service(IaaS).

2.4. Comprehensive Comparison between HPC Systems

In past Gabriel Mateescu *et. al.*, compared 9 key attributes of HPC, Grid, and Cloud. And also Ian Foster *et. al.*, Hameed Hussain *et. al.*, and Naidila Sadashiv *et. al.,* compared Cluster, Grid, and Cloud with several criteria in different perspectives. Table 1 depicts the summarized attributes among HPC systems after the arrangement of overlapping attributes.

2.5. Limitations of Past Researches

The researches described in the above section provide comprehensive perspectives to understand features of HPC systems at a glance. But these researches have several limitations. The First, these researches mostly focused on technical factors and didn’t describe application models precisely though application scientists consider not only technical factors but also other factors such as cost and benefits and so. The second, these studies compared factors only qualitatively hence user could not decide which infrastructure is fit to their research simply. The third, these researches compared them with each own criteria without unified standards. Consequently these researches addressed the attributes of infrastructure regardless to computation scientists as real users.

3. Analytic Hierarchy Process

Analytic Hierarchy Process developed by Thomas L. Saaty is a structure technique to derive several evaluation factors, and then deal with pair-wise comparison between them for structuring and analyzing decisions.13 AHP is practically used in a wide variety of domains such as government, business, education, and so on for decision making. Generally a decision making has the problem to find a best alternative in mutually incompatible criteria, incomplete information and limited resources.

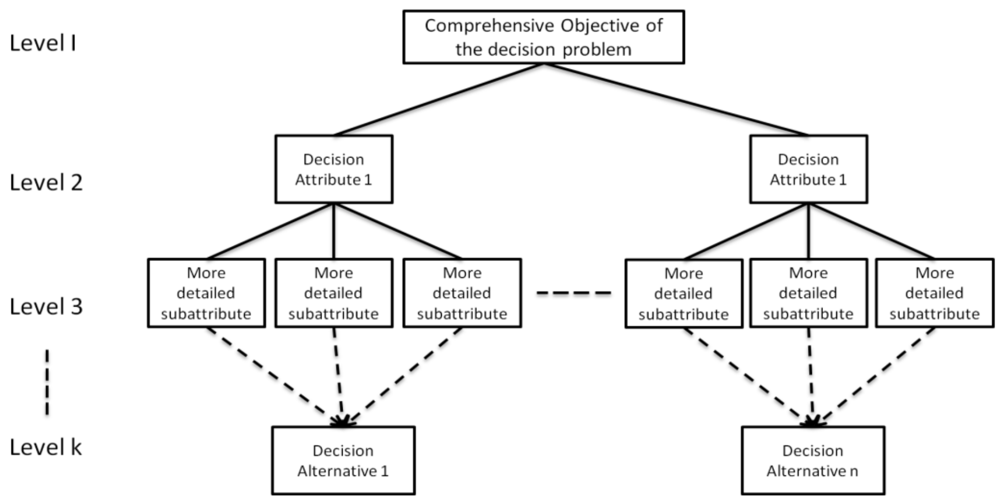


Figure 1. AHP Standard Hierarchy

AHP provides the capability to deal with the order of priority of several alternatives in various criteria. In AHP first, a pairwise comparison matrix should be made up. This matrix is consist of values which indicate that each attribute at the lower level under the upper level is better than other attributes at the same lower level through evaluating weights of each attribute at lower levels based on attributes at upper levels. And one normalized priority vector using the eigenvalue method is derived from each level of the hierarchy in this matrix. Finally one comprehensive priority vector of total hierarchy is derived, this vector shows relative priorities of alternatives at the lowest level to achiev*e.g.*eneral objective of the decision problem at the root level. Figure 1 depicts the standard hierarchy of AHP.

Based on the theory background described above, when AHP is applied to decision making problems practically the procedure of four steps is executed as follows.

Step 1 : set up a decision hierarchy through classifying the decision making problem as a hierarchy of mutually related decision attributes

Step 2 : collect judgment data via a mutually comparison between decision attributes

Step 3 : estimate relative weights of decision attributes using the eigenvalue method

Step 4 : summarize relative weights of decision attributes to calculate a comprehensive ranking of evaluated alternatives as the following equation(1).

$$C[1,k] = \prod_{i=2}^{k} B_i \qquad (1)$$

$C[1,k]$: a comprehensive weight of the kth decision attribute with regard to the first level.

B_i : $n_{i-1} \cdot n_i$ matrix including rows that are the structural elements of estimated vector

n_i : the number of attributes at *i*th level

And one of AHP features is to be able to analyze the sensitivity corresponding to changes of data related with a decision making problem. This makes it enable to examine how priorities of alternatives change corresponding to weight changes. There are five sensitivity analysis as follows : Performance Sensitivity, Dynamic Sensitivity, Gradient Sensitivity, Two Dimensional Plot, Weighted Difference Sensitivity.

4. Decision Support Model

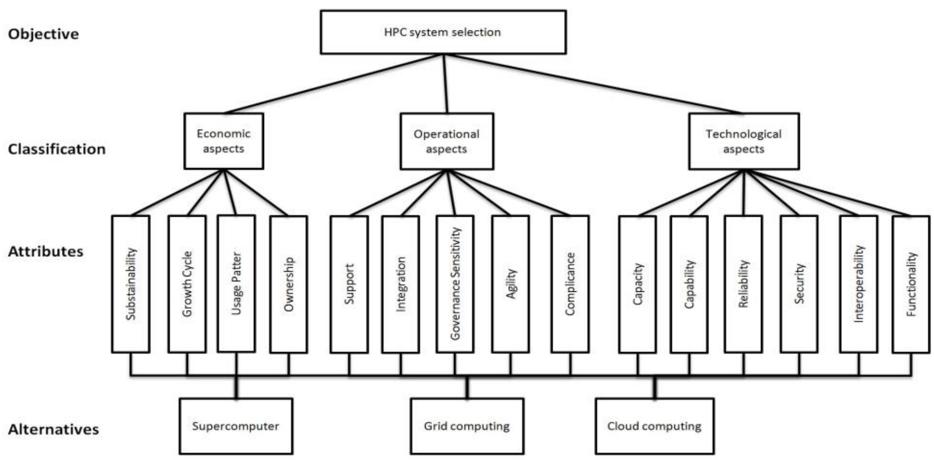


Figure 2. The Proposed Decision Support Model

For developing a decision support model, first we examine past researches described in the above section and derive decision criteria for the fitness of a HPC system selection. Basically the fitness of a HPC system selection is classified as an economic aspect, technological aspect, and operational aspect. In the economic aspect the benefit and cost to use a HPC system are considered. And In the technological aspect the performance and function that users need when they conduct their researches are considered. Also in the operational aspect when they use a HPC system additional elements such as administration, technical support, the size of their research are considered. Figure 2 depicts our proposed model to select a HPC system using AHP method.

Each attribute has the 9 scale degrees and each degree has the meaning as follows : 1 - Equal Importance, 3 - Moderate importance, 5 - Strong importance, 7 - Very strong importance, 9 - Extreme importance. And 2,4,6,8 values are the in-between value. And the fitness of a HPC selection is the product of evaluated values of attributes and weight of each attribute.

5. Summary and Future Works

5.1. Summary

Generally a supercomputer is defined as a high-end computer at the frontline of current processing capacity, by speed of calculation and storage capacity. At present the systems in the Top 500 supercomputer list released twice a year by Top500.org are designated, and they are considered as the most powerful general purpose computing systems that are in common use for high end applications around the world.

5.2. Future Works

In this paper we developed the decision support model but didn’t apply the model to read computational scientists using HPC systems. Thus we didn’t confirm that proposed decision attributes and subattributes were useful or should be refined by users in the field practically.

For validating our proposed model, we will contact several computation scientists using HPC systems and ask them to apply the model, and then take their feedback of criteria we selected. And also we will refine decision attributes and detailed subattributes in the model and find new decision attributes, too. This will make the model robust. Finally the cost of HPC systems is one of the most importance attribute which users consider thus we will study the cost analysis of HPC systems, and then unify or add the cost analysis model to the proposed model.

Acknowledgments

This work was supported by the program of the global hub for Experiment Data of Basic Science, 2015 funded by the NRF (N-15-NM-CR01-S01).

References

[1] K. Hwang, Editor, “Tutorial on Supercomputers : Design and Applications”, IEEE Computer Society, Los Alamitos, **(2003)**.

[2] J. P. Riganati and P.B. Schneck, “Supercomputing”, Computer, vol. 17, no. 10, **(1984)**, pp. 97-110.

[3] T. Hey, C.Scott and M.Surridge, “Simulation and Modelling Application on MPP Systems”, Proceedings of the 1994 EUROSIM Conference on Massively Parallel Processing Applications and Development, Delft, The Netherlands, **(1994)** June 21-23.

[4] H. Hussain, S. U. Rehman Malik, A.Hameed, S. Ullah Khan, G.Bickler, N.Min-Allah, M. Bilal Qureshi, L. Zhang, W.Yongji, N. Ghani, J. Kolodziej, A. Y. Zomaya, C.-Z. Xu, P.Balaji, A.Vishnu, F.Pinel, J. E. Pecero, D. Kliazovich, P. Bouvry, H.Li, L.Wang, D.Chen and A. Rayes, “A survey on resource allocation in high performance distributed computing systems”, Parallel Computing, vol. 39, no. 11, **(2013)**, pp. 709-736.

[5] G. L. Valentini, W. Lassonde, S.U. Khan, N. Min-Allah, S.A. Madani, J. Li, L. Zhang, L. Wang, N. Ghani, J. Kolodziej, H. Li, A.Y. Zomaya, C.-Z. Xu, P. Balaji, A. Vishnu, F. Pinel, J.E. Pecero, D. Kliazovich and P. Bouvry, "An Overview of Energy Efficiency Techniques in Cluster Computing Systems", Cluster Computing, vol. 16, no. 1, **(2013)**, pp. 3-15.

[6] N. Sadashiv and S. M Dilip Kumar, "Cluster, grid and cloud computing: A detailed comparison", Proceedings of the 6th International Conference Computer Science & Education, Singpore, **(2011)** August 3-5.

[7] D. Chen, L. Wang, X. Wu, J. Chen, S.U. Khan, J. Kolodziej, M. Tian, F. Huang and W. Liu, "Hybrid modelling and simulation of huge crowd over a hierarchical Grid architecture", Futur*e. g.*eneration Computer Systems, vol. 29, no. 5, **(2013)**, pp. 1309-1317.

[8] J. Kolodziej and S.U. Khan, "Multi-level hierarchic genetic-based scheduling of independent jobs in dynamic heterogeneous grid environment", Information Sciences, vol. 214, **(2012)**, pp. 1-19.

[9] Y. Amir, B. Awerbuch, A. Barak, S. Borgstrom and A. Keren, "An opportunity cost approach for job assignment in a scalable computing cluster", , IEEE Transactions on Parallel and Distributed, vol. 11, no. 7, **(2000)**, pp. 760-768.

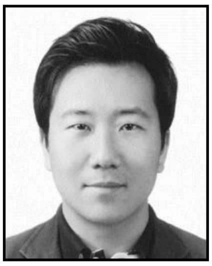
[10] I. Foster, Y.Zhao, I.Raicu and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared", Proceedings IEE*E. G.*rid Computing Environments, Austin, USA, **(2008)** November 12-16.

[11] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Futur*e. g.*eneration Computer Systems, vol. 25, no. 5-6, **(1999)**, pp. 757-768.

[12] R. Wolski, , N.T. Spring and J. Hayes, "The network weather service: a distributed resource performance forecasting service for metacomputing", Futur*e. g.*eneration Computer Systems, vol. 15, no. 6, **(2009)**, pp. 599-616.

[13] T. L. Saaty, Editor, "The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation", Mcgraw-Hill, Texas, **(1980)**.

Authors



Namgyu Kim is the senior researcher in Korea Institute of Science and Technology Information(KISTI) in South Korea. He obtained his bachelor's degree and master's degree in Computer Science and Engineering at Chung-Ang University in South Korea. He had participated lots of R&D projects including the national e-Science project, the AMGA development project, and project for construction of data Grid center on KEK Belle data and LIGO *etc.*At present he is in the department of Policy Research in KISTI.

Hyoung Woo Park is the director of national R&E network (KREONET) center and the Principal Researcher in KISTI Supercomuting center in south Korea. He obtained his Ph. D. in Computer Networks at SungKyunKwan University in south Korea. He had participated lots of R&D projects including the construction of national R&D network (KREONET), the implementation of National Grid computing Infrastructure, and the project for construction of peta-scale science data Grid center for th*e. g.*lobal collaboration researches on CERN LHC data, KEK Belle data *etc.*

Sudong Park is the principal researcher in Institute for Basic Science(IBS) in South Korea. He obtained his bachelor's degree, master's degree, and Ph. D. in Industrial Engineering at SungKyunKwan University in South Korea. Before IBS, he was the Research Fellow at Korea Institute of S&T Evaluation and Planning(KISTEP) in South Korea.

Hyungjin Lee, is the senior researcher in Korea Institute of Science and Technology Information(KISTI) in South Korea. He obtained his bachelor's degree and master's degree in Industrial Engineering, and his Ph. D. in Management of Technology at Sungkyunkwan University in South Korea. At present he is the executive secretary of National Institute of Supercomputing and Networking and interested in making a plan and strategy of national supercomputing.