Semantic Resource Classification Using Statistical Analysis for Application Characteristics in Intercloud Environment

Younsun Ahn Yoonhee Kim
Dept. of Computer Science
Sookmyung Women's University
Seoul, Korea
{ahnysun, yulan}@sookmyung.ac.kr

***Abstract*— Scientists gain benefits from scalable resource provisioning on-demand, and various computing environments by using cloud computing resources for their applications. However, many cloud computing providers offer their cloud resources according to their own rules. The descriptions of various cloud resources also differ for each vendor. Subsequently, it becomes difficult to find suitable cloud resources given the characteristics of an application. Scientists therefore, select resources used in previous experiments or best performance resources without considering the characteristics of their applications. There is the need to standardize notations to support simple selection of cloud resources without the constraints of providers. Intercloud can use cloud resources without considering cloud providers in hybrid cloud environments. Intercloud project has been studied for interoperability and creates mOSAIC ontology to conceptualize various resources. However, the mOSAIC ontology attributes are limited when considering characteristics of an application. We propose a semantic engine to provide semantic cloud resource services in intercloud environment. We define a rule of categorized resource description with reference to mOSAIC ontology attributes which includes added factors needed to represent application characteristics. We also develop a semantic engine which can choose semantically similar cloud resources using statistical analysis while considering the characteristics of an application. The semantic engine can also classify resources dynamically according to application specifications.**

Keywords*— *hybrid clouds; application characteristics; resoruce classification; semantic cloud

I. INTRODUCTION

Cloud Computing provides scalable resource provisioning on-demand, as well as various computing environments, using a pay-as-you-go method. Scientists, therefore, can take advantage of such mechanisms by using cloud computing resources for their computational applications. Despite the variety of resources available, scientists try to execute their application with specific resources which were used in previous experiments or they tend to select the best performance resources because it is difficult to find suitable cloud resources given the characteristics of some applications in hybrid cloud environment.

There are many cloud computing providers which offer their services according to their own rules. Thus it is hard to select the right resources from various cloud providers because descriptions of various cloud resources are different for each corporation. With these differences, it is hard to ensure interoperability among them. There is the need to standardize notations of cloud resources in order to easily choose cloud resources without constraints of providers.

Intercloud environment can use cloud resources without considering cloud providers in hybrid cloud environment. Intercloud project [1] works with associations and other institutes for standardization and has been studied for interoperability. Intercloud computing enables interconnection of multiple cloud venders' infrastructures. Intercloud can offer to cloud customers, diversified cloud environments in terms of providers.

Intercloud project creates mOSAIC [2] ontology which enables users to use various cloud resources. Representing hybrid cloud environment through mOSAIC ontology enables to conceptualize semantic attributes among various cloud resources. However, the mOSAIC ontology attributes are limited in taking into account the characteristics of an application.

Therefore, we propose a semantic engine to provide semantic cloud resource services in intercloud environment. We define a rule of categorized resource description which refers to mOSAIC ontology attributes and add factors which are needed to represent application characteristics. A semantic engine which considers the characteristics of an application is developed to semantically select similar cloud resources using statistical analysis, that is, cluster analysis [3]. To use any type of existing resource in the intercloud environment, we classify cloud resources and make each belong to a particular category according to the factors and characteristics of applications. Our semantic engine can also classify cloud resources dynamically according to application specifications.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A3007866)

The rest of the paper is organized into sections as follows; we introduce an overview of related works in Section 2. Section 3 explains semantic cloud architecture, Section 4 contains descriptions of a semantic engine algorithm, and in Section 5 experiment results are discussed. Finally, we conclude the paper and discuss future work in the final section.

II. RELATED WORK

Specifications of resources that cloud providers provide are totally different for each resource, yet cloud standardized classifications do not exist. Hence, to solve the problem, various studies have been conducted. Rimal [4] and Ankita Sharma [5] define classifications for specification of cloud services. They provide taxonomy of various cloud services as well as services of virtual resources. Study of cloud classification is limited to demonstrating the properties of several different resources of the various cloud providers.

Some papers proposed cloud resource management techniques using sematic method in a variety of ways such as using statistical techniques and using ontology. There are progressive studies on selection of cloud services involving a cloud ontology and calculations of similarity between concepts in ontology [5], and [6]. Several studies in active progress have classified cloud resources using a variety of statistical techniques [8], [9], and [10]. However, these studies also lack details on resource specifications for multiple cloud providers, so it is difficult to apply semantic techniques on the cloud services of various cloud venders.

In order to offer cloud resources to users in hybrid clouds without any constraints of cloud service venders. Intercloud project is studied for standardization and interoperability. Studies of resource selections relating to intercloud refer to intercloud project [1]. One of the intercloud project proposed mOSAIC ontology based on various vender's taxonomy and constructed mOSAIC ontology in order to conceptualize common property of cloud resources in hybrid cloud environment. The mOSAIC ontology provides interoperability between various cloud providers, integrates cloud environment to use cloud resources, and provides sematic searching among various cloud resources. However, the mOSAIC ontology attributes are limited in considering the characteristics of an application.

In this paper, we propose semantic engine that semantically selects similar cloud resources using cluster analysis which considers the characteristics of an application for user's convenience.

III. SEMANTIC CLOUD SERVICE ARCHETECTURE

This section gives an overview of semantic cloud service architecture. Semantic cloud service offers consumers appropriate select and search cloud resources for certain applications considering characteristics of that application. The aim of semantic cloud service is to provide services that suit an application's characteristics in hybrid cloud environment.

We propose the semantic cloud service architecture. The architecture comprises of resource specification which provides standardized cloud resource specifications of various venders', and application specification which represents characteristics of an application and critical factors for determining similarity. Resource specification enables users to choose various cloud service providers' virtual resources without difficulty.

Based on nine components, we make a rule to normalize cloud resource description, so that cloud resources in hybrid cloud environments can have common specifications for similar components. The nine components including provider's name, resource's name, memory size, physical machine's specification which is CPU, number of cores, and CPU flops, network bandwidth, number of vCPU, storage, and cost, were defined by referring to mOSAIC ontology components. We describe one hundred and seven cloud resources, supplied by Amazon EC2 [11], Microsoft Window Azure [12], Openstack [13], Rackspace [14], while applying a rule of resource description. With our proposed system, we can provide interoperability between various cloud providers such as Amazon EC2, Microsoft Window Azure, Openstack, Rackspace. The examples of resource description are shown in table 1.

TABLE 1. THE EXAMPLES OF RESOURCE DESCRIPTION

Resource Name	Provider	Memory size (GB)	CPU	CPU Number of cores	CPU Flops property	Network Bandwidth Property	vCPU	Storage (GB)	Cost
t2.micro	Amazon EC2	1	Intel Xeon Family	10	2.5	10	1	30	₩20
r3.large	Amazon EC2	15	Intel Xeon E 5-2670 v2	10	2.5	10	2	32	₩200
A4	Microsoft Azure	15	intel xeon E5-2660	4	2.2	10	8	2048	₩800
A6	Microsoft Azure	30	intel xeon E5-2660	4	2.2	10	4	999	₩1000
Standard3	KT Ucloud	2	intel Xeon L5640 2.29	6	2.27	10	2	100	₩74
Standard5	KT Ucloud	4	intel Xeon L5640 2.31	6	2.27	10	4	100	₩116
I/O1-120	rackspace	120	Intel Xeon Family	6	2.27	10	32	40	₩3,840
Memory1-240	rackspace	240	Intel Xeon Family	6	2.27	10	32	20	₩2,880
m1.small	Openstack	2	Intel Xeon Family	8	2.2	10	1	20	₩ -

Semantic engine, shown in the middle of the architecture performs tasks to classify similar cloud resources before providing semantic cloud services. Semantic engine calculates similarity between cloud resources using a statistical analysis that is a cluster analysis, and is based on specification of resources and an application. When the calculation occurs, semantic engine considers weights of resources depending on application specifications. Semantic Engine makes groups, which have similar resource information, with results from comparing resources. Therefore, according to application specifications, similar cloud resources are classified dynamically as groups.

Semantic cloud service facilitates selecting and searching for sufficient resources by checking groups with similar resource specifications. Semantic cloud service provides users with appropriate cloud resources considering characteristics of an application in hybrid cloud environment regardless of the cloud service provider.

IV. SEMANTIC ENGINE ALGORITHM

In this section, we describe how semantic engine works and its possible effects. We propose a semantic engine algorithm for executing that application in hybrid cloud environment considering characteristics of an application. We measure semantic similarity between cloud resources using cluster analysis. Cluster analysis can group sets of cloud resources such that cloud resources in the same group are more semantically similar to each other than to those in other groups. Semantic engine algorithm includes 3 steps: 1. Prerequisite step, 2. Calculating step, and 3. Grouping step. First, semantic engine takes into account weighting factors in resource specification to satisfy characteristics of an application. It also brings information of cloud resources which are organized based on nine components.

In Calculating step, all resources without considering providers, are compared to each other respectively in order to calculate similarity between them. In this step, when the semantic engine compares resources having resource factors, it assigns weights to special resource factors needed to consider characteristics of an application. Semantic engine attains values, which means similarity between resources, through the comparisons. The values are represented in a two-dimensional matrix in order to show comparisons to each resource separately.

Finally, the grouping step separates cloud resources according to similarity. Semantic engine classifies resources using values of a comparison. Cloud resources in the same group are more semantically similar to each other than to those in other groups. If characteristics of an application are different, then Semantic engine takes into account weights which are suitable for the applications. Therefore, Semantic engine obtains dynamic results according to characteristics of an application.

V. EXPERIMENT

We experiment a proposed algorithm on hybrid clouds. We execute autodock application [15] with Amazon EC2, Window Azure, Openstack. First, we classify the cloud platforms as groups considering characteristics of the autodock application. Then, we execute autodock on some cloud resources which are sorted by the semantic engine.

A. Assumptions

In the field of applied Computer Science, computational biology is the use of computer science and statistics to solve problems in biology. We use an application of autodock which is one of the applications in computational biology. Much of computational biology is involved with the analysis of molecular structure. An application of autodock performs automated docking of a flexible ligand to a protein. An application of autodock is a modeling method which predicts optimal ways of interaction of a flexible small ligand with a strict molecular target. An execution of an autodock application takes relatively short time per-task and requires a large number of docking used in high-performance computing. Therefore, an application of autodock needs to regard memory size and the number of vCPU, so, we assign weights to memory size and the number of vCPU according to characteristics of the application in computational biology.

B. Result of cluster analysis

Fig. 1 shows a result of calculating similarity. Semantic engine deals with two weighting factors as mentioned before with regards to characteristics of autodock application.

First, we represent one hundred and seven cloud resources offered by Amazon EC2, Microsoft Window Azure, Openstack, Rackspace applied a defined rule of resource description.

Semantic engine calculates similarity with descriptions by comparing cloud resources to each other. In the graph represented in fig.1, each cloud resource is represented by one symbol. To reflect the application characteristics, semantic engine calculates the similarity of cloud resources for the application of autodock.

This result is customized to the autodock application as a result of considering factors such as memory size and number of vCPU.

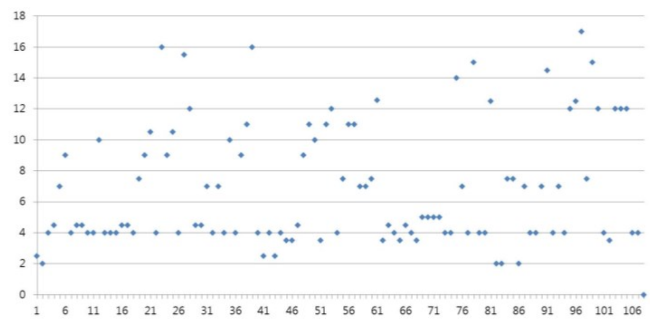

Fig. 1. The result of calculating similarity

Fig. 2 displays a result of cluster analysis. Semantic engine categorizes cloud resources as clusters using the result of calculating similarity. Cloud resources are grouped into different circles as seen in fig.1. The greater the degree of similarity of weighting factors of the cloud resources, the more likely they are to be classified in one cluster.

In the result, a total of 10 clusters were created as shown in fig.2. The number of cloud resources in each cluster is

different for each. When a user decides to execute their application in cloud resources, the user chooses one of the clusters. Users choose a cluster according to their desire or according to the difference between the standard deviation for running stable executions.

Fig. 2. The result of cluster analysis

In each cluster, cloud resources' specification is similar so users can select similar cloud resources in one of the clusters created using cluster analysis. In each cluster, the cloud resources are provided by various venders, therefore, a user cannot know who the provider is and what the provider's own cloud resource descriptions are. Consequently, the user easily picks semantically similar resources in intercloud environment. Users can also predict the result of executing their application in each cluster through the result of cluster analysis.

C. *Result of executing application after cluster analysis*

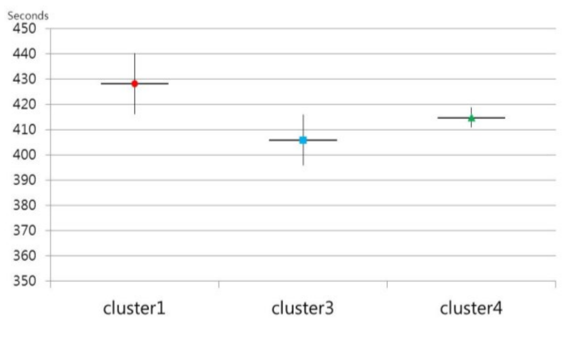

Fig. 3. The result of experiments

We choose some cloud resources following the result of cluster analysis and we execute an autodock application in the cloud resources to confirm that cloud resources are well grouped by using cluster analysis.

We select t2.micro (Amazon EC2) and m1.small (Openstack) for cluster1 as shown in fig. 2, m3.2xlarge (Amazon EC2), c4.xlarge (Amazon EC2) for cluster2, A2 (Window Azure), and A4 (Window Azure) for cluster3, and r3.large (Amazon EC2) and A6 (Window Azure) for cluster4 as shown as fig. 2. The specifications for each chosen resources are represented in table 1. We deploy the application in each chosen cloud resource and examine the execution time.

Fig.3 displays the average execution time and standard deviation of each cluster. In cluster1, the execution time of autodock application spans an average of 428.23 seconds with standard deviation of 23.98 points. In cluster3, the average execution time of autodock application is measured as 405.88 seconds with standard deviation of 20.05. Lastly, cluster 4 shows an average execution time of 405.88 seconds with standard deviation value of 20.05.

Through the experiments, we ascertain that our semantic engine can classify well. Execution time of autodock application tends to be similar in each cluster. Semantic engine sorts semantically similar cloud resources with consideration to the application characteristics.

VI. Conclusion and Future Work

We propose a semantic engine to offer semantic cloud resource services in intercloud environment. We define a rule in order to represent cloud resource descriptions. Semantic engine provides users with the option to semantically select similar cloud resources using cluster analysis while considering application characteristics. We experiment using an application of autodock and give the result of grouping similar cloud resources according to characteristics of the autodock application. Semantic engine can classify cloud resources dynamically according to application characteristics.

In the future, we plan to extend semantic engine and simulate experiments.

References

[1] Intercloud project, http://www.intercloudtestbed.org/

[2] Moscato, F, Aversa, R, Di Martino, B, Fortis, T, Munteanu, V, "An Analysis of mOSAIC ontology for Cloud Resources annotation", in Computer Science and Information Systems (FedCSIS) 2011 Federated Conference on IEEE, pp. 973-980, September. 2011.

[3] Eisen, Michael B, Spellman, Brown, Botstein, "Cluster analysis and display of genome-wide expression patterns", Proceedings of the National Academy of Sciences, 95(25), pp. 14863-14868, 1998.

[4] RIMAL, Bhaskar Prasad, CHOI, Eunmi, LUMB, Ian. "A taxonomy and survey of cloud computing systems", in 5th IEEE International Joint Conference on INC, IMS and IDC (NCM'09), pp. 44-51, August. 2009.

[5] Ankita Sharma, Sonia Vatta, "Cloud Computing Taxonomy and Architecture", International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 5, pp. 1410-1417, May. 2013

[6] Li Liu, Xiaofen Yao, Liangjuan Qin, Miao Zhang, "Ontology-based Service Matching in Cloud Computing", in Fuzzy Systems, 2014 IEEE International Conference on IEEE, pp. 2544-2550, July. 2014.

[7] Rodríguez-García, Miguel Ángel, et al. "Ontology-based annotation and retrieval of services in the cloud.", Knowledge-Based Systems 56, pp.15-25, January. 2014.

[8] Qingtao Wu, Min Cui, Mingchuan Zhang, Ruijuan Zheng, Ying Lou, "A Cloud Service Resource Classification Strategy Based on Feature Similarity", Journal of Networks, Volume 9, No 11, pp.2987-2993, November. 2014.

[9] Md Whaiduzzaman, Abdullah Gani, Nor Badrul Anuar, Muhammad Shiraz, Mohammad Nazmul Haque, Israat Tanzeena Haque, "Cloud Service Selection Using Multicriteria Decision Analysis", The Scientific World Journal, Volume 2014, February. 2014.

[10] Kourtesis, Dimitrios, Jose María Alvarez-Rodríguez, and Iraklis Paraskakis. "Semantic-based QoS management in cloud systems: Current status and future challenges." Future Generation Computer Systems, Volume 32, pp. 307-323, March. 2014.

[11] Amazon Web Service, http://aws.amazon.com/

[12] Microsoft Windows Azure, http://www.windowsazure.com/

[13] Openstack, http://www.openstack.org/

[14] Rackspace, http:// www.rackspace.com/

[15] Ocana, K. ; Benza, S. ; De Oliveira, D. ; Dias, J. ; Mattoso, M., "Exploring Large Scale Receptor-Ligand Pairs in Molecular Docking Workflows in HPC Clouds", Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International, pp. 536-545