A Service-Oriented Qos-Assured and Multi-Agent Cloud Computing Architecture

Bu-Qing Cao[1,2,3,*], Bing Li[1,2], and Qi-Ming Xia[1,2]

[1] State Key Laboratory of Software Engineering, Wuhan University, Wuhan, 430072, China
[2] School of Computer, Wuhan University, Wuhan, 430072, China
[3] School of Computer Science and Engineering,
Hunan University of Science and technology, Xiangtan, 411201, China
cao6990050@163.com, bingli@whu.edu.cn

Abstract. The essence of Cloud Computing is to provide services by network. As far as user are concerned, resources in the "Cloud" can be extended indefinitely at any time, acquired at any time, used on-demand, and pay-per-use. Combined with SOA and Multi-Agent technology, this paper propose a new Service-Oriented QOS-Assured cloud computing architecture which include physical device and virtual resource layer, cloud service provision layer, cloud service management and multi-agent layer to support QOS-Assured cloud service provision and request. At the same time, based on the proposed service-oriented cloud computing architecture, realization process of cloud service is simplified described.

Keywords: cloud computing; SOA; Multi-Agent; QOS-Assured; architecture.

1 Introduction

With the significant advances in Information and Communications Technology over the last half century, there is an increasingly perceived vision that computing will be the 5th utility one day (after water, electricity, gas, and telephony) [1].There are many technologies which enable cloud computing are still evolving and progressing, for example, Web2.0 and Service Oriented Computing [2].Now, IT companies are now talking about establishment environments of cloud computing, however, if each company will build its own cloud computing platform, this will result in isolated cloud. Therefore, it is especially important to design architecture for cloud computing. At present, different manufacturers offer different design schemes, which result in readers can't entirely understand cloud computing principles. Combined with SOA and Multi-Agent technology, this paper proposes a new Service-Oriented QOS-Assured and Multi-Agent cloud computing architecture which includes physical device and virtual resource layer, cloud service provision layer, cloud service management and Multi-Agent layer to support QOS-Assured cloud service provision and request.

* Corresponding author. Tel.: +86-027-61304188 , fax: +86-027-68754590.

M.G. Jaatun, G. Zhao, and C. Rong (Eds.): CloudCom 2009, LNCS 5931, pp. 644–649, 2009.

2 Service-Oriented Qos-Assured and Multi-Agent Cloud Computing Architecture

This section firstly introduces SOA and Multi-agent technology; secondly, cloud service Qos model is established according to the characteristic of cloud service consumer and provider; finally, a Service-Oriented QOS-Assured and Multi-Agent cloud computing architecture is designed to support QOS-Assured cloud service provision and request.

2.1 SOA and Multi-Agent Technology

There are advantages for service management and architect-driven concept in the SOA [3-6].Currently, cloud computing technology has hardly any service management and architect-driven concept. Therefore, many companies choose to wait-and-see attitude rather than rush to adopt it. So, the idea of service management and architect-driven can be applied to cloud computing. By this, cloud computing can be seen an extension which SOA provides resources to "cloud", such as, IaaS, PaaS, SaaS, and its key is to determine which cloud services, information and processes on the cloud is the best candidate, and which cloud services should be abstracted in the existing or emerging SOA.

Software agent is a software entity which runs continuous and independent in a given environment, usually combined other agents with solving problem [7]. Multi-Agent system has been increasingly attracted to researchers in various fields, particularly in the network environment, agent can be used to complete complex task by communicating with many resources and task publishers. Cloud computing refers to both the applications delivered as services over internet and hardware and systems software in the datacenters[8], and it provides a variety of resources, such as network, storage, computing resources to users adopted by IaaS, PaaS, SaaS and other forms of service. These resources are vast, heterogeneous, distributed; it is very important how to provide them to users with high-quality, validity. Described by the above, agent can be used to complete complex task by communicating with many resources and task publishers. So, it can be used in service-oriented cloud computing architecture to support QOS-Assured cloud service provision and request.

2.2 Cloud Service Qos Model

There have many related research work on QOS, but QOS mentioned in many articles mentioned only relate to consumers. Taking into account strong background resource process and service provision capabilities, this paper considers all related QOS attributes of cloud service consumer and cloud service providers. As far as cloud service providers, cloud service Qos provided by the physical device and virtual resources layer mainly focus on data center's performance, reliability, stability; cloud service Qos provided by Iaas likely emphasize on response time, resource utilization, and prices, and so on. As far as cloud service consumers, they are very important, such as, response time, price, availability, reliability, reputation, and they can also be provided by the service provider. Thus, considering Qos of cloud services providers and consumers, the most common attributes of Qos will be illustrated as follows and other attributes can be extended according to different service form.

Definition 1 CloudServiceQOS$_{responsetime}$(S). It represents the interval from the requirement sending of cloud service consumers to cloud service implements competition, which is calculated as follows:

$$\text{CloudServiceQOS}_{responsetime}(S)=\text{Time}_{transfers}(S)+\text{Time}_{run}(S). \quad (1)$$

Among this, $\text{Time}_{transfers}(S)$ on behalf of the transmission time form requirement sending to results return and it can be gained by cloud service monitor; $\text{Time}_{run}(S)$ represent cloud service implements time and it also can be obtained by cloud service monitor.

Definition 2 CloudServiceQOS$_{cost}$(S). It represents fees paid when customer use service provided by cloud service provider, that is, pay-per-use, and it can be realized by cloud service meterage.

Definition 3 CloudServiceQOS$_{availability}$(S). It represents that the probability of cloud services can be accessed, which is calculated as follows:

$$\text{CloudServiceQOS}_{availability}(S) = A/N. \quad (2)$$

Among this, N express the request times that consumer want to use cloud service S during a certain period of time; A express the accessible times of cloud service S.

Definition 4 CloudServiceQOS$_{reliability}$ (S). It show the capacity that cloud service accurately implements its function and the times of validation and invalidation can be acquired by cloud service monitor, which is calculated as follows:

$$\text{CloudServiceQOS}_{reliability}(S)=R/M. \quad (3)$$

Among this,R express the times of called and successful implements of the cloud service S; M on behalf of the total called times of the cloud service S.

Definition 5 CloudServiceQOS$_{reputation}$(S). It expresses the creditability of cloud services. Reputation can be seen as the sum of subjective customer's rating and objective QoS advertising messages credibility (CoWS) (based on the Bayes learning theory),in order to reduce the impact of malicious rating[9], which is calculated as follows:

$$\text{CloudServiceQOSreputation}(S)=\alpha\times\left(\sum_{i=1}^{n}\text{Rating}\right)/n+\beta\times\text{coWS}. \quad (4)$$

Thus, according to above given definition of cloud services related Qos, an integrated Qos model of cloud service S can be expressed as follows:

$$\text{CloudServiceQOS}(S)=W1\ |\text{CloudServiceQOS}_{responsetime}(S)|+W2\ |\text{CloudServiceQOS}_{cost}(S)|+W3\ \text{CloudServiceQOS}_{availability}(S)+W4\ \text{CloudServiceQOS}_{reliability}(S)+W5\ \text{CloudServiceQOS}_{reputation}(S) \quad (5)$$

Here , $W_i \in [0,1]$, $\sum_{i=1}^{5} W_i = 1$, W_i express the weight of corresponding Qos i and its value can be set according to user preferences, for example, user preferences is lower prices and faster response time in an air ticket booking service, thus, the values of W_1 and W_2 can be set to bigger. |CloudServiceQOSresponsetime (S) | expressed QOS attributes dimensionless or normalized process. Specific dimensionless process method does not belong to the scope of this study. This model sustain the extension of

QOS attributes, that is can add or remove QOS attributes according to specific situation, to support QOS-assured cloud service acquired at any time, used on-demand, pay-per-use and extended indefinitely.

2.3 Service-Oriented QOS-Assured and Multi-Agent Cloud Computing Architecture

Figure1 shows a Service-Oriented QOS-Assured and Multi-Agent cloud computing architecture which includes physical device and virtual resource layer, cloud service provision layer, cloud service management and Multi-Agent layer, to support QOS-Assured cloud service provision and request.

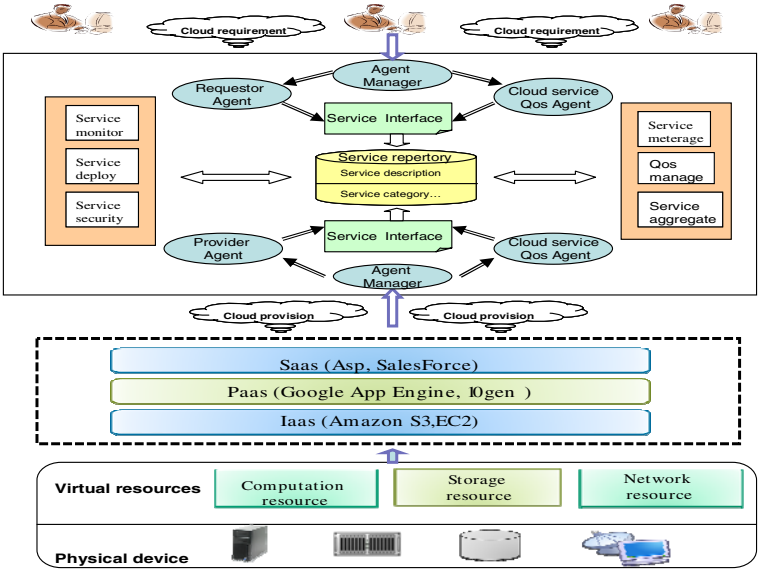

Fig. 1. Service-Oriented and Multi-Agent Cloud Computing Architecture

1) Physical Device and Virtual Resource Layer
Physical resources is all kinds of physical equipment which support upper services of cloud computing, such as a large number of servers in data center, network equipment, storage equipment and so on. Cloud computing is a shared-resource computing method by the form of virtualization. Here, physical resources can be converted into various resources, such as computing resources, storage resources, network resources by virtualization technology, then they can be connected together to form a flexible, unified resources pool in order to dynamically allocated to different applications and service requirement, thereby improve resources utilization rate.

2) Cloud Service Provision Layer
Cloud service provision layer can provide some forms of services by functions composition provided by physical device and virtual resource layer. The forms of service

that cloud computing provides today may be broken down into managed services, SaaS, Web services, utility computing, and PaaS. Figure 1 shows a concentrated services view of cloud computing, including Iaas, Paas, and Saas, which provide IT service capabilities to users.

3) Cloud Service Manager and Multi-Agent Layer

Cloud service manager and multi-agent layer mostly manages a variety of services provided by cloud service provision layer and finds QOS-assured cloud service in service repertory according to user's cloud service requirement. As shown in Figure 1, cloud services management which includes service repertory, service interface, service aggregate, service monitor, service deploy, service meterage, service security, QOS manage. Among them, service repertory similar to UDDI in the SOA, which includes service description, service category, and so on. Service description represents service functional and non-functional information, such as service names, response time, and so on; service category represents service type provided by cloud service provision layer, such as IaaS, Saas, Paas; service interface represents access interface according to services types, for example, Remote Desktop for IaaS, API for PaaS, web services for SaaS; service aggregate represents that new services can be created based on existing services; service monitor represents monitor and alarm according to health status of the services and automatically correct abnormity state of services; service deploy represents automatically deploy and configure specific services examples according to users requirements; service meterage represents cloud services, similar to water, electricity and gas, which are available and pay on-demand by the user; service security represents provide authorization control, trust management, auditing, consistency check for cloud services; QOS manage represents that manage cloud service Qos model which select, calculate and dimensionless process Qos attributes and described in section2.2, at the same time, sustain the extension of QOS attributes, that is can add or remove QOS attributes according to specific situation, to support QOS-assured cloud service acquired at any time, used on-demand, pay-per-use and extended indefinitely, and return the best expected cloud service to user.

Cloud multi-agent management which includes cloud service requestor agent, cloud service QOS agent, cloud service provider agent and agent manager, and it is mainly support QOS-assured cloud service provision and request. Cloud services requester is no longer direct visit cloud service but commit to requestor agent, who mainly collects feedback information of request and submits cloud service request based on QOS. Similarly, Cloud services provider is no longer direct publish services but commit to provider agent, who mainly collects cloud service use information and publishes cloud service. Cloud service QOS agent primarily submits cloud services QOS information. Agent manager primarily manages various managers, such as new, recovery operation.

Thus, the process of cloud service provision can be described: First of all, Cloud service provision layer can provide various cloud service to agent manager and then establish provider agent and cloud service QOS agent. Secondly, service function information, such as service description, service category, which can be standardized by service interface, will be submitted to service repertory by provider agent. At the same time, service QOS information, such as, response time, cost, which can be standardized by service interface, will be submitted to service repertory by cloud service QOS agent. The process of cloud service request is the same to the process of cloud service provision.

4 Conclusion and Future Work

This paper proposes a new Service-Oriented QOS-Assured and Multi-Agent cloud computing architecture to support QOS-Assured cloud service provision and request. There are research challenges in the future for cloud computing that make use of open standards and architecture to allow different clouds for interoperability. Next research work will focus on cloud services interoperability research.

Acknowledgement

This work is supported by National Basic Research Program (973) of China under grant No. 2007CB310801, National High Technology Research and Development Program (863) of China under grant No. 2006AA04Z156, National Natural Science Foundation of China under grant No. 60873083 and 60803025, Research Fund for the Doctoral Program of Higher Education of China under grant No. 20070486065,National Science Foundation of Hubei Province for Distinguished Young Scholars under grant No.2008CDB351 and National Science Foundation of Hubei Province under grant No.2008ABA379.

References

1. Buyya, R., et al.: Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. Future Generation Computer Systems 25(6), 599–616 (2009)
2. Wang, L., von Laszewski, G.: Cloud Computing: a Perspective Study
3. Toward a service-oriented analysis and design methodology for software product lines, http://www.ibm.com/developerworks/webservices/library/ar-soaspl/
4. Vouk, M.A.: Cloud Computing-Issues, Research and Implementations. In: ITI 2008, Cavtat, Croatia (2008)
5. de Leusse, P.: Secure & Rapid composition of infrastructure services in the Cloud. In: The Second International Conference on Sensor Technologies and Applications, Cap Esterel, France (August 2008)
6. van der Burg, S., Dolstra, E.: Software Deployment in a Dynamic Cloud: From Device to Service Orientation in a Hospital Environment. In: ICSE CLOUD 2009, Vancouver, Canada, May 23 (2009)
7. Adshaw, J.M.: An Introduction to Software Agents. In: Bradshaw, J.M. (ed.) Software Agents, pp. 3–46. AAAI Press, Menlo Park (1997)
8. Armbrust, M.: Above the Clouds: A Berkeley View of Cloud Computing, EECS Department, University of California, Berkeley (2009)
9. Li, Z., Su, S., Yang, F.C.: WSrep: A novel reputation model for web services selection. In: Nguyen, N.T., Grzech, A., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2007. LNCS (LNAI), vol. 4496, pp. 199–208. Springer, Heidelberg (2007)