INCREASING SLA BASED ACCESS CONTROL IN CLOUD COMPUTING

D. MADHAVI
Research Scholar, OPJSU,
Churu, Rajasthan, India

Dr. SATEESH KUMAR NAGINENI
Associate professor, Dept of CSE,
OPJSU, Churu, Rajasthan, India

ABSTRACT
Cloud computing is the most advanced technology in the real world environment and provides flexible and convenient possibilities for users to utilize available services. Resource provisioning to the satisfaction of user requirements becomes the most challenging task in the heterogeneous cloud environment. Proper admission control algorithms need to be proposed for better resource provisioning with improved user satisfaction level. In this research, Knowledge based Service Level Agreement (SLA) aware admission controlled scheduling and resource allocation is proposed which makes use of machine learning algorithms namely Support Vector Machine (SVM) and Artificial Neural Network (ANN) for better admission control. It seeks to study the knowledge of resource status information by using machine learning algorithms in the training phase. Based on these strategies, admission control would be done in the testing phase which would lead to efficient and better resource provisioning.

Keywords: *Service Level Agreement, Artificial Neural Network (ANN), Cloud computing.*

1. INTRODUCTION
This section presents a comprehensive introduction to cloud computing, types of services provisioned by the cloud computing environment and so on. It also provides a detailed overview of the different scheduling and admission control policies.

1.1 An Overview of Cloud Computing:
Cloud computing is a popular field in the today's world which gives ease of accessing several types of services to people in a flexible manner. It is already entering into the next phase of evolutionary growth for supporting people, industries, and organizations by providing the most optimal services. The most vital sector of the world known as Information Technology (IT) prefers the integration of cloud services into their domains. The cloud services allow the IT sector to efficiently share their content in/out of the organization to develop their industries (Irena Bojanova et al. 2014).

Cloud computing is determined as the maintaining and delivering the most valuable services such as software, application, and storage services to users as per their specifications in the best possible practical modes. It reduces the burden of users from storing the data in their host and managing them periodically. Cloud service providers (CSP) are the ones who set up the cloud servers and designate the expected services to the users. (Office of the privacy commissioner of Canada, Factsheet 2009)

1.1.1 Popularity of Reason Behind the Cloud Computing:
Cloud computing services are aggressive and favoured by most of the public, organizations, enterprises and government institutions. The prevalence of cloud computing services is becoming increased due to various reasons and factors. Some of the main reasons are mention as below (Office of the privacy commissioner of Canada, Factsheet 2009):

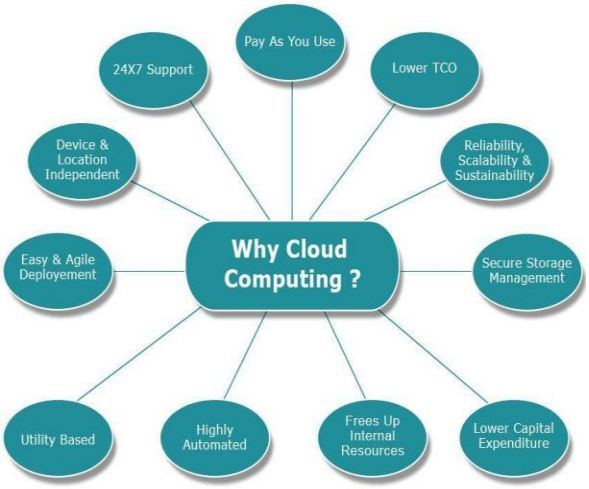
Decreased service usage cost with more versatility.

More scalability of resources, providing the dynamic user specifications well.

More reliability regarding provisioning resources in a continuously.

Enhanced satisfaction of users regarding receiving better services that satisfy their requirements.

A graphic description indicating the causes behind the prevalence of cloud service is provided in Figure 1.1.



(Source: www. todaydatacenter.com)
Figure 1.1 Reasons behind the popularity of cloud services

From this figure, it is clear that scalability, reliability and low estimate cost are the most significant reasons for the growing requirement for cloud computing services.

1.1.2 Cloud Service Models

Cloud service models are classified into three types depends on the type of service provided (VMware white paper 2012).
Infrastructure as a Service (IaaS)
Software as a Service (SaaS)
Platform as a Service (PaaS)

IaaS preparations the infrastructure-oriented components to users like storage, virtual machines (VM), firewalls, load balancers etc. One of the best examples of IaaS service providers is Amazon web services (Eugene Gorelik 2013).

SaaS gives the software services to users. This service allows users to obtain complete access to the software application. And SaaS types of service providers assure that users would get upgraded software to perform their tasks in a better way. The Social media platforms are better examples of the SaaS service providers (Eugene Gorelik 2013).

PaaS service provides comprehensiveness to users project working experience by offering complete pre-built application platforms. And PaaS provides an Application Programming Interface (API) to the clients to access their PaaS services flexibly. Google app engine is an excellent representative for PaaS service providers (Eugene Gorelik 2013).
A pictorial illustration of various types of cloud computing models regarding the cloud service provisioning functionality given in Figure 1.2

Figure 1.2 Cloud service models

This figure also incorporates the several services that are delivered by the various cloud service models.

1.1.3 Cloud Distribution Models

Cloud computing demands an Internet connection for providing the services to the users. Several types of organizations and industries are set up the cloud environment and allocate the resources to users based on their requirement. The cloud computing models can be classified into four categories regarding the deployment (Chakradhara Rao et al. 2013).
Public Cloud
Private Cloud
Hybrid Cloud
Community Cloud

Public Cloud is the one the most formed externally to implement various services like web application, storage spaces etc. to the users to allow them to utilize the contents. Public cloud is not specific to one specified group, and it can employ by any person who has located anywhere in the world. A shadow is termed as a public cloud only when the services are provided as openly available services for public use. The structure of the public and private cloud is substantially similar, but just the security tolerances vary from each other and differ for the services, applications and accessible resources for public use.

Private Cloud is a national service provider that functions only for specific kind of people to implement the services. It is formed by some organization or industry to facilitate content distribution to the group of people who relate to them. Private cloud is exclusively used and operated by the single private organization either maintained or hosted internally or by third-parties. The higher degree of engagement to virtualizes business environment is required to initiate a private cloud. The primary issue is security threat and necessity of protection against severe vulnerabilities.

Community Cloud is designed for some specific set of organizations having a similar condition and plans for performing their responsibilities. It is natural for some groups and to share their contents through the population cloud network. It is a collaborative effort where the support of various organizations but within a particular community is combining regarding general issues like security, compliance, jurisdiction, etc.

Hybrid cloud is an integration of two or more cloud service models collectively. And It is determined as the integration of hardware of various servers concurrently to enable them to share contents through virtualization environment. Hybrid cloud allows private and public cloud services to combine and offer the resources.

A pictorial representation of the cloud deployment models is given in Figure 1.3.

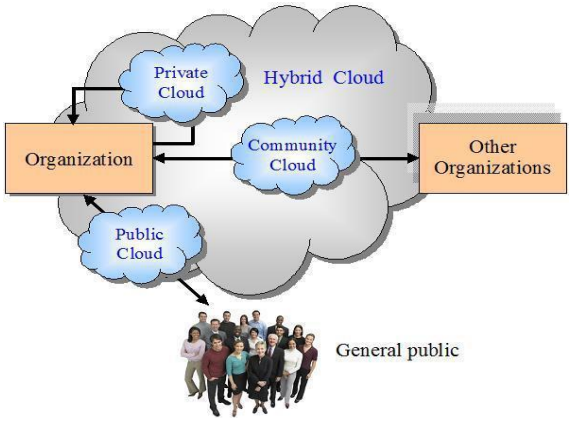


Figure 1.3 Cloud deployment models

2. LITERATURE REVIEW

This chapter deals with the study of various research works that have been previously conducted for establishing admission control methods in cloud, cloud task scheduling and resource allocation with due consideration for the various QoS constraints.

2.1 ADMISSION CONTROL IN CLOUD COMPUTING

Marcus Carvalho et al. (2015) proposed a prediction-based admission control for IaaS clouds with multiple service classes which have different pricing and service level objective (SLO) guarantees. The approach maximizes the request admission rates and also fulfills the availability of SLOs defined for each class. SLO guarantee is provided by using the proposed model by instantiating it with different forecasting methods for limiting the admission of VM requests for different classes, and trace-based simulations using data from production systems.

Shahdi-Pashaki et al. (2015) introduced the group based technology for the optimal resource allocation process which would combine the similar set of tasks in terms of QoS parameters and the resource allocation would be done for those similar set of tasks. After grouping, resource allocation for those

large size tasks is done in the optimal manner by using the cuckoo search optimization algorithm which will search for the most optimal resource that satisfies the QoS requirements. Cuckoo search optimization approach allocates the tasks into the resources in a round robin manner. The computational results of this work proved to provide better results than existing algorithms.

Ron C Chiang & Howie Huang (2014) proposed a novel Task and Resource Allocation Control framework (TRACON) which performs a better and efficient task of scheduling with the consideration of the data dependency exists between the set of similar tasks. Resource consumption details are also considered for better allocation of resources with the consideration of various factors. This work designs the interference prediction module for predicting the resource consumption details in order to ensure the promising delivery of resources in a continuous manner. Interference aware scheduler is used to allocate the resources that match with the user task requirements with the consideration of the data interference between those tasks. Task and resource monitor will gather the resource and task execution status dynamically to find and avoid the overlapping of resources which occurs in run time.

Sudip Misra et al. (2014) implemented a Learning Automata (LA) - based QoS (LAQ) framework to design a framework for better service allocation with the assured QoS constraints. This is achieved by continuously monitoring the resource and task allocation status information, so the resources can be allocated in a better way. Also it enables user to get the required resources from cloud service provider for the period of time they requested. This resource allocation can also be adjusted easily based on the user‟s on-demand requests in run time. Learning automata concept is adapted for implementing this framework which would allow cloud service providers to move from one state to another state with the penalty probability, thus the requirement of both service providers and the cloud users can be fulfilled.

Alok Gautam Kumbhare et al. (2013) introduced the reactive resource allocation approach which can allocate the resources into the cloud tasks in terms of different scheduling mechanism. This work adapts the genetic algorithm for the optimal resource allocation in terms of allocation of optimal resources. This overall mechanism adapts the bin packing problem for ensuring the assured delivery of resources to cloud users in terms of guaranteeing QoS constraints. This mechanism also ensures the optimal resource provisioning for the better allocation of tasks into the cloud resources. The overall research of this work assures that this methodology can allocate the resources with the satisfaction of various QoS constraints.

2.2 GROUPING AWARE RESOURCE ALLOCATION

Linlin Wu et al. (2012) proposed a SLA based admission control mechanism to improve the resource allocation in the SaaS service providers‟ environment with due consideration of objectives such as improved user satisfaction and profit. In this mechanism, the new incoming requests are stored in the waiting queue and will be made to wait until the jobs allocated earlier are completed. The advantage in this mechanism is that before allocating and scheduling the jobs, the broker would calculate the expected profit of the user submitted jobs. Based on the profit value, the task scheduling would be performed. The drawback in this mechanism is the presence of error in the estimated service time that can degrade the overall performance.

Kleopatra Konstanteli et al. (2012) presented a novel admission control test for deciding the admission of services into an elastic cloud by obtaining the minimum allocation

for each of the components that contain the services. The mechanism focuses on hosting elastic services which means that the required resources may increase or decrease dynamically based on the dynamically varying amount of users and patterns of the demands. The estimation of the optimal number of allocations depends on the admission control test that incorporates the business rules in terms of trust, co-efficiency and cost. It also takes account of affinity rules and the components that comprise the services.

Josep L Berral et al. (2010) proposed the energy aware scheduling using machine learning algorithms to provide a flexible environment for the service providers to scale down their energy consumption. The energy consumption is often related to factors such QoS parameters like power consumption levels, CPU loads, and SLA timings and so on. These need to be estimated well for the better prediction of the resources in terms of achieving enhanced scale down control of energy efficiency. The overall research of this work was evaluated in the practical heterogeneous cloud environment by considering the QoS parameters. This research work leads to better resource handling capability considering scaling down or scaling up of the resources.

Hong-Ha Nguyen et al. (2009) solved the problem of grid sites scheduling process with the satisfaction of QoS parameter values. This optimal grid resource allocation process achieves better utilization by allocating the tasks into resources that can process well with the promising feature of scalability. This approach works well by making use of reserved computing resources and the light path scheme which is utilized well by introducing the joint scheduling approach. This approach finds the logical relationship between the available resources and the start time of tasks, and the resource collision can thus be avoided successfully.

Kevin Power et al. (2008) introduced the novel mechanism for performing resource allocation process for the real time services by grouping the resources. This grouping of resources is done to satisfy the heterogeneous resource requisition of real time applications. This work gathers the user requirements which would be grouped together based on the QoS constraints. The group of users with similar QoS constraints would be assigned to the shared resource that satisfies QoS constraints. After this initial allocation, data flow rate will be calculated in terms of QoS constraints based on which resource allocation would be redone to achieve it. This group based resource allocation strategy achieves promised optimal resource allocation scenario with the improved satisfaction of user QoS constraints.

Bo An et al. (2007) improved the resource allocation process in the multi agent system with incomplete information. This was done by augmenting the resource allocation scheme with the optimization strategy which attempted to achieve better resource allocation strategy by resubmitting bid details to the resource providers. This resubmission of bid values optimizes the dealing stage, and thus the optimal resource allocation can thus be achieved. The overall experimental result of this work provides not better result than the allocation mechanism with complete information, but however it can optimize the dealing strategy by resubmitting the bid information to the trading agents, so that they can optimize the resource allocation strategy.

Gaurav Pandey et al. (2005) introduced novel active Support Vector (SV) learning algorithms that are able to adapt the combinations of random and query learning. These mechanisms can perform scheduling with the satisfaction of users QoS constraints. The proposed algorithms provide better results in terms of achieving fast and robust convergence. Random SVM would assume that all the predicted patterns

are independent from each other by choosing the correct predictable answer. In query SVM, output would be predicted by finding the solution that is most near convergence to the hyper plane which depicts the QoS factors. This approach can reach most optimal solution by following the convergence solution.

Mardente et al. (2004) designed a novel entity called domain controller involving cloud tasks and cloud service providers to perform the admission control process. Domain controller will perform admission control tasks by analysing the factor called the traffic flow. The domain controller approach utilizes traffic flow detection algorithms to detect the level of traffic flow in resources. It will prevent the task from submission to the cloud resource that has more traffic flow where there might be a possibility of task execution failure.

3. SUMMARY

This chapter has highlighted the findings from the literature review on the importance of minimizing the execution time, failure rate and increasing profit and improving QoS constraints. This section also has given the detailed overview of the different techniques/methodologies that can be used for obtaining the scheduling with increased profit.

4. CONCLUSION

Cloud resource provisioning is a highly sensitive task which needs to be done with greater concern for increasing the user satisfaction level as well as the profit level of the cloud service providers. Admission control is one of most popular technologies which can perform optimized resource provisioning in the presence of heterogeneous environment with varying user requests. Better and effective admission control task can lead to improved utilization of cloud resources with increased profit for cloud service providers. This is achieved by introducing various research methodologies that attempt to allocate the resources with the concern for increased profit and reduced makespan.

In the first research called machine learning based admission control, SVM and ANN algorithms are used for performing effective scheduling. Both the proposed approaches attempt to control the admission control process by allocating the tasks to the resources that can provide more Return on Investment (ROI). This is done by learning the profit and the non-profit parameters of the cloud resources and the tasks in an enhanced manner.

REFERENCE

[1] Aarti Singh, Dimple Juneja & Manisha Malhotra 2015, „A novel agent based autonomous and service composition framework for cost optimization of resource provisioning in cloud computing", Journal of King Saud University – Computer and Information Sciences.

[2] Alok Gautam Kumbhare, Yogesh Simmhan, Marc Frincu & Viktor, K. Prasanna 2015, „Reactive Resource Provisioning Heuristics for Dynamic Dataflows on Cloud Infrastructure", IEEE Transactions On Cloud Computing, vol. 3, no. 2.

[3] Amit Agarwal & Saloni Jain 2014, „Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment", International Journal of Computer Trends and Technology (IJCTT) – vol. 9, no. 7.

[4] Andrei Tchernykh, Uwe Schwiegelsohn, Vassil Alexandrov & El-ghazaliTalbi 2015, „Towards Understanding Uncertainty in Cloud Computing Resource Provisioning", ICCS 2015 International Conference On Computational Science, vol. 51, pp. 1772-1781.

[5] Ang, TF, Ng, WK & Ling, TC A Bandwidth-Aware Job Grouping-Based Scheduling on Grid Environment. Information Technology Journal, vol. 8, no. 3, pp. 372-377.

[6] Bahman Javadi, Jemal Abawajy & Rajkumar Buyya 2012, „Failure-aware resource provisioning for hybrid Cloud infrastructure", Elsevier Journal of Parallel and Distributed Computing, vol. 72, Issue

10, pp. 1318-1331, DOI: /10.1016/j.jpdc.2012.06.012.

[7] Balaji Palanisamy, Aameek Singh, Ling Liu & Bhushan Jain 2011, „Purlieus: Locality-aware Resource Allocation for MapReduce in a Cloud", Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis.

[8] Bassem Wanis, Nancy Samaan & Ahmed Karmouch 2015, „Efficient Modeling and Demand Allocation for Differentiated Cloud Virtual-Network as-a Service Offerings", IEEE Transactions on Cloud Computing.

[9] Bhavani, BH & Guruprasad, HS 2014, „Resource Provisioning Techniques in Cloud Computing Environment: A Survey", International Journal of Research in Computer and Communication Technology, vol. 3, Issue 3.