Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.783 – 788

A REVIEW ON SERVICE LEVEL AGREEMENT AND SLA BASED RESOURCE PROVISIONING IN CLOUD COMPUTING

P.Tamil Selvi[1], B.Satheeskumar[2]
[1]Department of Computer Science, Bishop Heber College, India
[2]Department of Computer Science, Bishop Heber College, India
[1] thayarachel@gmail.com; [2] satbhc@yahoo.com

Abstract*— Cloud computing have paved a way to access any type of resources such as hardware, software and infrastructure over the internet and peculiarly on-demand. These resources in cloud are either owned or managed by Cloud Service Provider (CSP) and cloud vendor. SLA is an agreement signed between the end user and the Cloud Service Provider or cloud vendor to ensure the quality of service that is being offered by them. The SLA mainly targets to gratify the specified Service Level Objectives (SLO) to intensify both customer satisfaction levels together with an increase in the profit rate from the provider side and this is achieved by using some distinct admission control and scheduling algorithm that go well with SLA.

Keywords*— "Cloud", "SLA", "Resource Provisioning", "Scheduling", "Admission Control"

I. INTRODUCTION

Cloud computing have paved a way to access a shared pool of resources over the internet irrespective of time and location. One of the bedrock advantages of cloud computing is "pay – as-you-go" model and "on-demand" provisioning. Though utility computing and grid computing follows metered and utility service model but they have not obtained a great popularity unlike cloud computing because of the lack of virtualization and multi-tenancy properties. Cloud computing has incorporated virtualization, multi-tenancy, on-demand provisioning of resources, metered services, resource pooling etc. To put all together the U.S National Institute of Standard and Technology (NIST) have proposed a formal definition for cloud computing: "*Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]* This definition provided by U.S NIST comprehends the architecture of the cloud, the deployment model and the service delivery model of the cloud.

A. The five main properties of cloud computing are as follows:

"On demand self-service" A consumer who is in demand for any computing resources such as storage, CPU time, software, infrastructure for some time can utilize the same indeed without making any prior apprising with the resource provider. There is no need to make any reservation for the resources that a user wants to access.

"*Wide network access"* Cloud computing is not limited to any constrains like platform and application dependency rather the services are delivered over the internet to the end user regardless of the network design and the platform the user is using from their site.

"Resource pooling" The computing resources are pooled in the cloud and are managed by the cloud service provider. Through "multi-tenancy" and "virtualization" the resources that are needed by various consumer and user are provided.
"Rapid Elasticity" The resources provided to the end user can be scaled-up and scaled- down (i.e.) the resources can be randomly accessed if needed beyond what they have and randomly released if the resource is more than that what is required
"*Measured Services"* The resources that are purchased and utilized by the cloud consumer is measured and billed based on the usage. That's why cloud computing is often termed as "Pay as you go".

II. SERVICE LEVEL AGREEMENT

In cloud environment, the services are provided by cloud service provider or by cloud vendors. To access the resource the end user and the cloud service provider has to make a covenant with each other.

A. What is SLA

Service Level Agreement (SLA) is a legal agreement signed between the customer and the service provider that ensures the Quality of Service (QoS) parameters like data storage, availability of the system CPU, and network. Hence the SLA should describe,

- The roster of services provided by the service provider and the explanation of each service.
- A system of measurement to quantify the services delivered by the provider is as declared and a system of auditing to overseer the services.
- Liability of the provider and consumer.
- Description about the solution if the SLA is violated either by provider or consumer.
- A narration about the changes that may happen in SLA in course of time.
  The above objectives are considered as Service Level Objectives (SLO) that may also include QoS parameters.[6]

B. Types of SLA

SLA can be classified based on the interaction between the consumer and the provider. The classification of SLA may be,

i. Provider Pre-Delineated
ii. Negotiated SLA

The provider defined SLA offers a general SLA model for the customer. At times the customer may have expected a specific quality of service demands which may not be covered in the provider pre-delineated SLA. In that circumstance, the customer seeks a negotiation with the provider about the quality of service that the customer demands. The service provider should handle certain mechanism to ensure both the quality of service to the customer and without any deterioration in their profit [7]. These issues are addressed by researchers using some SLA based resource provisioning management mechanism and negotiation strategies. [2][3]

C. Layered Architecture of cloud

There are three layers that plays chief role while providing resources. The *Infrastructure as a Service (IaaS) layer* is considered to be the resource provisioning model since any infrastructure resources like hardware, servers, networking components and storage facilities are offered by the service provider to the customer based on the on-demand needs of the consumer. The next layer is the *Platform as a Service (PaaS) layer* that serves the consumer with development tools, execution management services, solution stack and computing platform. The *Software as a Service (SaaS) layer* delivers the software application to the end user using the resources supplied in PaaS and Iaas layers.

D. Parties involved in SLA

The computing resources in the cloud are provided to the end user by means of various services like SaaS, Paas, Iaas, DaaS and NaaS etc. In order to access the resource the end user and the cloud provider or consumer signs a legal treaty known as SLA.

1) The parties involved in signing the SLA:
Cloud Service Provider (CSP): They actually own and manage the resource in cloud. The cloud consumer rent resources from CSP. These CSP provides services like storage, network and CPU time etc.
Cloud Vendors or Cloud Consumers: Cloud hosted software application which utilize the CSP services, are managed by the cloud vendors. They did not own the resource but lease the resources from the CSP. They are financially liable for the resource consumption.
End user: They represent the actual user who makes the request to access resource from the cloud consumer or cloud service provider and consumes the resource.

"The cloud service provider charge cloud consumer for renting resources to deploy their application, cloud consumers may charge their end user for processing their workloads(e.g Software as a service) or they may process the user requests for free (cloud-hosted business application) ." [4]

2) There are two types of SLA based on application hosting. They are:
Infrastructure SLA: These SLA are provided by the cloud infrastructure service provider to the cloud consumer. They manage and offer an assurance about the quality and availability of the infrastructure like server machine performance and their power, network speed and its connectivity, storage capacity and availability. The machines that are leased to the customer are isolated from each other.
Application SLA: These SLA are provided by the cloud consumer and the end user. They vouch for the quality of service for the applications and software that they have deployed over the infrastructure layer of the cloud. The cloud service provider allocates and releases the computing resources based on the demand among the co-hosted applications. Thus the cloud service provider is indirectly responsible to satisfy the customer's application service level objectives.

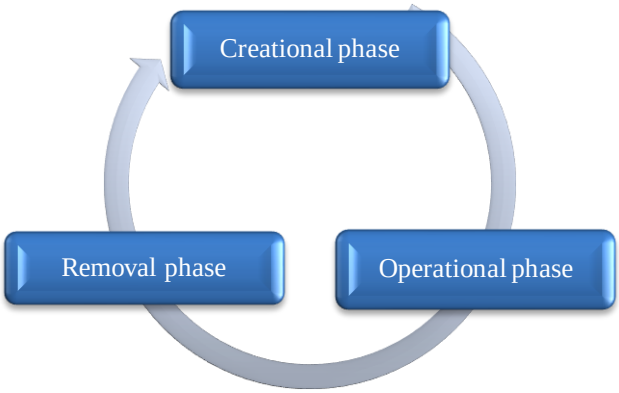

Fig. 1 Different Phases in SLA life cycle

At each level the participants involved in signing the SLA may vary. The infrastructure SLA is signed between the Cloud Service Provider and the cloud consumer or vendor. The Application SLA is signed between the cloud consumer and the end user.

E. Need for SLA

Enhanced customer satisfaction level: SLA helps to concentrate on the requirements and wishes of the customer and helps the service provider and the vendor to have a self-evaluation that the effort to achieve the SLO is on the exact direction.
Improved Service Quality: Key Performance Indicator (KPI) is the measured specification level that is associated with each item specified in SLA and this helps to specify whether the cloud consumer or the service provider is able to service the customer in a better way.
Improved relationship between two parties: The Service Level Objectives are QoS parameters specified in SLA. These parameters are monitored by the cloud consumer so that if any issue may arise it is easy to resolve. Moreover the rewards and the penalties if SLA is violated is specified in SLA so that through this the end user may get a bright idea how to utilize the resource without violating the objectives of the SLA.

F. Need for SLA

The three phases of SLA are as follows (e.g. Fig.1)

i. *Creation Phase*
ii. *Operation phase*
iii. *Removal phase*

The creation phase is the first phase, in which the cloud consumer or the end user seeks the good service provider who can satisfy their requirements and needs. If the cloud service provider matches the requirements of the consumer and the end user, they move to the next phase, the operational phase. This phase allows the consumer and the end user to have a read-only access to the SLA. If the consumer or the end user feels fine with the pre-defined SLA of the Service provider, then an agreement is

made and the resource is provisioned or else they move for negotiated SLA. After all the operation is over, the removal phase encounters. In the final Phase the SLA that is signed during the resource provisioning is de-signed and all other associated configuration is released. The three phases figures the principle stages only. The phases may have many sub-phases. For example the operational phase deals with establishing agreement, provisioning resources, monitoring and measuring service. It may also deal with the penalties that should be enforced if SLA is violated.

III.RELATED WORKS

Various researches have been carried on the topic SLA and risk management in the domain of cloud computing. In order to overcome the security issue that is affiliated with SLA in the cloud, researchers provide SLA management models and framework. There are three major SLA based factor that should be analysed such as risk factor that is associated with the service, the service cost and the response time. The following lists some of the research that has been made in the area of SLA and risk management in cloud computing. Alhamad et al [9] modelled various models in the area of SLA-based trusted model for cloud computing. Using this model the cloud consumers can evaluate the resources that are available in the cloud and can decide the resource that is more reliable to them. In addition to this, Alhamad et al [10] also proposed a SLA framework for cloud computing. This framework provides good criteria for discussing about the negotiation strategies between the cloud providers and the other participants like cloud consumer, cloud broker, or SLA's monitoring agent, thereby helping to build a good SLA in cloud. Hammadi and Hussain [11] proposed a monitoring framework for SLA. This framework helps the third party providers to monitor whether the specification of the SLA is being met by all the parties in real time. Reputation assessment module and transactional risk assessment module are the two modules that are used in this framework. These modules allow the consumer to take good decision by providing them with a real-time QoS assessment. Chi et al. [12] presented "SLA-tree" which is a data structure to make decisions based on SLA in cloud environment. It contains two data sets: One is the list of waiting queries to be executed and the other set is the SLA for each four query – which indicates the different queries profits to modify response time for each query. Jahyun Goo [13] proposed a framework to structure SLA in outsourcing the IT arrangements. The detailed description about measurement development and accurate statistical validations of the SLA is explained in this framework. This framework encompasses contractual factors of eleven SLA and three more sub-factors that are related with it. Hedwig et al [14] modeled a SLA design for enterprise information system. This framework deals with the state-of -art of different concept that is related with system management and it helps to balance the risk associated with the process cost. This design helps the IT leaders are able to understand a clear picture about the correlation between the process cost and the service quality. Zhang et al [15] presented a framework for information security risk management in cloud environment. This framework helps to recognize the critical areas in cloud and to identify the impact of threats and vulnerabilities in the cloud environment. Moreover it focuses on the possible measurements that should be taken to alleviate the risk. Several issues and challenges of SLA and risk management in cloud is presented by Morin et al [16]. To identify and quantify risks in cloud computing environments a framework as in [15] is used. Cloud Security Alliance (CSA) in its "Cloud security guide" [17] has insisted the cloud consumers to employ security departments so that the security requirements in the SLA can be enforced. The process of risk management in this paper uses FERMA standard [18]. In order to examine the relationship between the network availability and SLA specification the risk analysis is also been carried out.

IV. RESOURCE PROVISIONING BASED ON SLA

Resource provisioning mechanism deals with the discovering, allocation and monitoring of resources. When the user requests for the resource the service provider or the consumer search and discovers the resources and it is allocated. The infrastructure provider allocates the physical resources that are requested by the service provider. There is no direct interaction between the end user and the infrastructure provider. The resource is allocated based upon the SLA that ensures the quality of service like response time, performance, reliability and availability. While provisioning resources, the service provider should monitor the resources such that the resource is not over-provisioned or under-provisioned and it is the responsibility of the service provider to ensure that the terms specified in SLA is not violated.

Type of resource provisioning based upon the need of the application.

i. *Static provisioning*
ii. *Dynamic provisioning*

In static provisioning the resources are allocated based upon prediction. This type of provisioning technique works well if the demands of the application does not change in course of time. The consumer signs a contract and makes request for the service in advance, the provider will fix the resources before the service starts. A flat-fee on the monthly basis is charged for the consumer.

Dynamic provisioning of resources is allocated based upon the user request in on-demand. Allocation based upon prediction does not comport dynamic provisioning since the demand of the application may vary often. If the user is running out of resource then the service is added to them if a request is made by the user. If the user wants to release the resource if it is not needed then it is done so. Hence dynamic provisioning always charge the customer based on "pay-as-you-go" model. But dynamic provisioning model has so many challenges like Virtual Machine migration, new virtual machine instance creation, re-

allocation, re-scheduling the new request and power consumption. *Cloud bursting* is the term that refers to creation of hybrid cloud using dynamic provision technique. [5]

Factors involved in Resource provisioning.

Response time: The time taken to respond a task must be minimized. The algorithm should be designed in that model.

Cost of the service: The cost of using the cloud must be minimized from the end user perspective.

Revenue charges: The cloud provider aims at maximize so, from their aspect.

Fault tolerance: The algorithm should be designed in the way that it provides access even at the time of any node failure.

SLA Violation: The algorithm should concentrate on reducing SLA violation.

Power management: The technique that is used for virtual machine placement and virtual machine migration should also consider the lower power consumption.

V. ADMISSION CONTROL AND SCHEDULING ALGORITHM

Admission control:

When a new request arrives, admission control algorithm decides which request can be allowed to run in the application server at times of heavy loads. The benefit of using admission control algorithm is to monitor the incoming request and find a subset of the incoming request to be allowed to execute on the server when the system undergoes an overload situation that in turn increases the overall pay-off. From the study of literature the admission control mechanism is classified as Request and session based.

In *request based mechanism,* new request are rejected if the capacity of the server is low. This may sometimes do not serve important jobs or the subset of currently executing job and sometimes may honor unimportant jobs

In *Session based mechanism,* the new sessions are rejected until the longer session gets completed, so that none of the subset of the currently executing job request are not rejected rather they are admitted. The rejection of new request based upon the type of the user making the request and the nature of the request. For example, during overload situation, the high-priority jobs are accepted and the lower-priority jobs are rejected. The request that consumes more system resources are rejected and the request that consumes less resources are admitted. [8]

Scheduling Algorithm:

Scheduling algorithms distributes the load among the processor and thereby reducing the total task execution time. Scheduling mechanism help to maximize throughput and performance rate by scheduling the job to the best adaptable resources in accordance with the adaptable time. Job Scheduling in cloud may be classified into two groups. They are Batch (BMHA) and online mode heuristic algorithms. In BMHA, when job are arrived they are collected and queued in a set and the scheduling will start after some period of time. First come first serve Scheduling algorithm, Round Robin Scheduling algorithm, Min-Min algorithm, Max-Min algorithm are some of the examples of BMHA.

In Online Mode heuristic mechanism, when jobs arrives the system they are scheduled, rather they are not pooled and scheduled. Most fit task scheduling algorithm is an example for online mode heuristic algorithm.

REFERENCES

[1] P.Mell and T.Grance, "Draft NIST working definition of cloud computing – v15," 21.Aug 2009, 2009.

[2] Zukernine, F., and Martin, P. (2011). An Adaptive and Intelligent SLA Negotiation System for Web Services. IEEE Transactions of Service Computing, 4(1), (pp. 31-43).

[3] Shell. M., Comuzzi, M., and Pernici, B. (2007). An Architecture for Flexible Web), 2012 IEEE 5th International conference on (pp. 360-367) Service QoS Negotiation. Proceedings of the 1st IEEE International Enterprise Distributed Object Computing (EDOC) Conference, Maryland, USA.

[4] Sherif Sakr and Anna Liu, "SLA-Based and Consumer- Centric Dynamic provisioning for Cloud Databases" Cloud Computing (CLOUD), 2012 IEEE 5th International conference on (pp. 360-367)

[5] Bhavani B H1 and H S Guruprasad, Resource Provisioning Techniques in Cloud Computing Environment: A Survey. International Journal of Research in Computer and Communication Technology, Vol 3, Issue 3, March- 2014

[6]" Review and summary of cloud service level agreement" "Cloud Computing Use Cases Whitepaper" Version 4.0 , 2010

[7] Linlin Wu " SLA- based Resource Provisioning for Management of Cloud-based Software-as-a-Service Applications"

[8] Sumit Bose, Anjaneyulu Pasala, Dheepak Ramanujam A, Sridhar Murthy and Ganesan Malaiyandisamy, "SLA management in Cloud computing: A service provider perspective

[9] M. Alhamad, T. Dillon, and E. Chang, "SLA-Based Trust Model for Cloud Computing," in *Network-Based Information Systems (NBiS), 2010 13th International Conference*, 14-16 Sept., pp. 321-324.

[10] M. Alhamad, T. Dillon, and E. Chang, "Conceptual SLA framework for cloud computing," in *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference*, 13-16 April 2010, pp. 606-610.

[11] Adil M. Hammadi and Omar Hussain, "A Framework for SLA Assurance in Cloud Computing," in *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference*, 26-29 March 2012, pp. 393-398.
[12] Yun Chi, Hyun Jin Moon, Hakan Hacigumus, and Junichi Tatemura, "SLA-tree: a framework for efficiently supporting SLA-based decisions in cloud computing," in *In Proceedings of the 14th International Conference on Extending Database Technology (EDBT/ICDT '11)*, New York, NY, USA, 2011, pp. 129-140.
[13] Jahyun Goo, "Structure of service level agreements (SLA) in IT outsourcing: The construct and its measurement," in *Information Systems Frontiers 12*, April 2010, pp. 185-205.
[14] M. Hedwig, S. Malkowski, and D. Neumann, "Risk-Aware Service Level Agreement Design for Enterprise Information Systems," in *System Science (HICSS), 2012 45th Hawaii International Conference* , Jan. 2012 , pp. 4552-4561.
[15] Xuan Zhang, N. Wuwong, Hao Li, and Xuejie Zhang, "Information Security Risk Management Framework for the Cloud Computing Environments," in *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference*, June 29 2010- July 1 2010, pp. 328-1334.
[16] J. Morin, J. Aubert, and B. Gateau, "Towards Cloud Computing SLA Risk Management: Issues and Challenges," in *System Science (HICSS), 2012 45th Hawaii International Conference*, 4-7 Jan. 2012, pp. 5509-5514.
[17] Cloud Security Alliance. (2011, October) Security Guidance for Critical Areas of Focus in Cloud Computing v3.0. [Online]. https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf
[18] FERMA. Risk Management Standard. [Online]. http://www.ferma.eu/riskmanagement/standards/risk-management-standard/

