Autonomic Cloud Computing: Research Perspective

Research · July 2015

DOI: 10.13140/RG.2.1.4453.4881

1 author:

Sukhpal Singh Gill
Queen Mary, University of London
116 PUBLICATIONS **2,917** CITATIONS

SEE PROFILE

Published Online: 2 July, 2015

Author's Viewpoint

Computer Science and Engineering Department,
Thapar University, Patiala, Punjab, India

Autonomic Cloud Computing: Research Perspective

Sukhpal Singh Gill
ssgill[at]thapar[dot]edu

Introduction

Cloud computing is an evolving utility computing mechanism in which cloud consumer can detect, choose and utilize the resources (infrastructure, software and platform) and provide service to user based on pay per use model as computing utilities. Current computing mechanism is effective, particular for medium and small cloud based companies, in which it permits easy and reliable access to cloud services like infrastructure, software and platform. Present cloud computing is almost similar to the existing models: cluster computing and grid computing. The important key technical features of cloud computing which includes autonomic service, rapid elasticity, end-to-end virtualization support, on-demand resource pooling and transparency in cloud billing. Further, non-technical features of cloud computing includes environment friendliness, little maintenance overhead, lower upfront costs, faster time to deployments, Service Level Agreement (SLA) and pay-as-you-go-model. In distributed computing environment, unpredictability of service is a fact, so same possible in cloud also. The success of next-generation Cloud Computing infrastructures will depend on how capably these infrastructures will discover and dynamically tolerate computing platforms, which meet randomly varying resource and service requirements of Cloud costumer applications.

Logically, based on QoS requirements such as scalability, high availability, trust and security, these applications will be characterized, identified in the so called SLAs. The current Cloud technology is not completely personalized to honor probable SLAs, though industrial and the academic, both the research groups are presenting increasing interest on problems of QoS assurance within the context of Cloud Computing. Broadly, an SLA needs a precise assessment of the characteristics of the required resources. Application services introduced in Clouds (e.g., Web applications, Web services) are frequently characterized by great load inconsistency; therefore, the amount of resources required to honor their SLAs may vary particularly over time. An important challenge for Cloud providers is to automate the management of virtual servers while keeping into account both high-level QoS requirements of hosted applications and resource supervision expenses. Cloud market mechanisms are consistently static and cannot react on dynamic variation of consumer desires. To respond to these issues, there is a requirement of an adaptive methodology for autonomically springing SLA patterns based on consumer requirements. The present research in Cloud SLA limits the capability of matching conformation metrics to acceptable. These metrics comprise statistical measures such as standard deviation that want to be computed from the expected and actual outcomes of services delivered to customer. Semantic Web technologies can be used to improve the descriptions and therefore increase the quality of these matches. Although, Cloud consumers do not have full supervisory control over the fundamental computing resources, they do require ensuring attributes such as quality, accessibility, trustworthiness, and performance of these resources when users have transferred their fundamental business functions onto their honored Cloud.

Published Online: 2 July, 2015

Analysis of SLA Deviation

SLA designates what you require from your consumers/service customers in order to provide the service specified. It needs assurance and support from both parties to provision and follow the contract in order for the SLA to work efficiently. In SLA, both the parties (Cloud Provider and Cloud Consumer) should have specified the possible deviations to achieve appropriate quality attributes. Assume that there are 8 different Cloud services being providing by Cloud Provider with the same SLA deviation (4%) for each service by taking availability as a quality attribute as shown in Table.

Table: Cloud SLA Deviations

Cloud Service	EA Without SLA Deviation	EA With SLA Deviation	AA	Actual SLA Deviation	SLA Violates
CS1	98%	94%	91%	7%	Yes
CS2	96%	92%	95%	9%	No
CS3	97%	93%	94%	3%	No
CS4	91%	87%	80%	11%	Yes
CS5	93%	89%	90%	3%	No
CS6	99%	95%	92%	7%	Yes
CS7	98%	94%	95%	3%	No
CS8	99.2%	95.2%	92.4%	6.8%	Yes

Table describes the Expected Availability (EA) without SLA deviation, Expected Availability (EA) with SLA deviation, Actual Availability (AA), Actual SLA deviation and whether the SLA violates for a particular service or not. The comparison of SLA deviations of 8 Cloud services is shown in Figure. The comparison of Actual deviation (vary) and Expected deviation (4%) of 8 Cloud services is shown in Figure. Usually, no Cloud provider is considering compensation because 85% resource providers do not actually provide penalty enforcement for SLA violation presently. There should be penalty delay cost or consumers' compensation if the Cloud provider misses the deadline. Moreover, it provides a risk transfer for IaaS providers, when the terms are desecrated by Cloud provider. Penalty delay cost is how much the service provider has to give concession to users for SLA violation. It is dependent on the penalty rate and penalty delay time period. The effect of inaccuracy could be reduced by two approaches: first, considering the penalty compensation clause in SLAs with IaaS provider and impose SLA violation; second, adding some slack time during scheduling for avoiding risk.

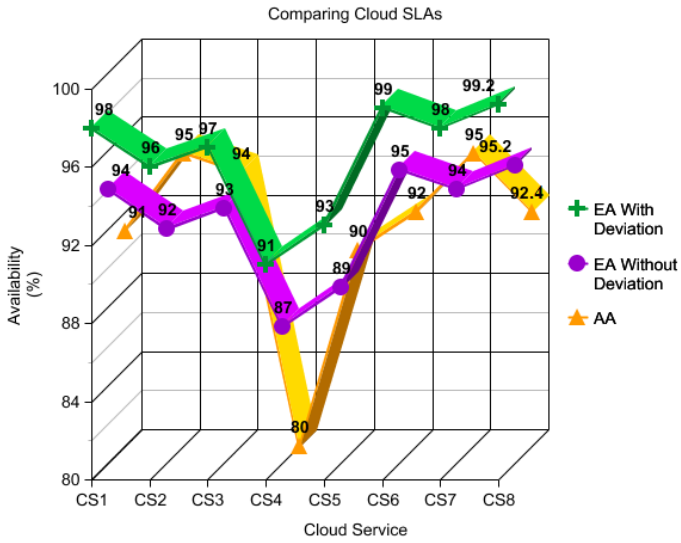

Figure: Comparison of SLA deviations

Published Online: 2 July, 2015

In other words, it is vital for users to acquire assurances from suppliers on service provisions. Usually, these are delivered through SLAs discussed between the providers and customers. The very first problem is the description of SLA terms in such a way that has a suitable level of granularity, namely the compromises between accuracy and complexity, so that they can ensure most of the user hopes and is comparatively simple to be prejudiced, certified, calculated, and imposed by the resource provisioning mechanism on the Cloud. In addition, different Cloud service models (IaaS, PaaS, and SaaS) will need to express different SLA meta disclaimers. This also increases a number of implementation issues for the Cloud providers. Moreover, innovative SLA mechanisms require to continuously integrate consumer response and customization features into the SLA assessment framework. As the Cloud service models develop and become omnipresent, it increases the probability of clarifying the way the services are provisioned and managed. It, therefore, permits the providers to address the different requirements of their customers. In this perspective, SLAs appear as a significant characteristic which subsequently serves as the establishment for the predictable quality level of the services made available to customer by the providers. Nonetheless, the collection of the recommended SLAs by providers (with marginal overlaps), has directed to manifold different definitions of Cloud SLAs. Moreover, confusions exist on what is (if there is) the difference between SLAs and agreement, what is the marginal quality, what are the terms involved in each one of these documents and if and how are these associated.

References

[1] Sukhpal Singh, and Inderveer Chana. "Q-aware: Quality of Service based Cloud Resource Provisioning." Computers & Electrical Engineering - Journal - Elsevier. [http://dx.doi.org/10.1016/j.compeleceng.2015.02.003]

[2] Sukhpal Singh, and Inderveer Chana. "QRSF: QoS-aware resource scheduling framework in cloud computing." The Journal of Supercomputing, Vol. 71, no. 1, pp: 241-292, 2015.

[3] Sukhpal Singh, and Inderveer Chana. "Cloud Based Development Issues: A Methodical Analysis." International Journal of Cloud Computing and Services Science (IJ-CLOSER) vol.2, no. 1 , pp. 73-84,2012.

[4] Sukhpal Singh, and Inderveer Chana. "Advance Billing and Metering Architecture for Infrastructure as a Service." International Journal of Cloud Computing and Services Science (IJ-CLOSER) vol.2, no. 2 pp.123-133,2013.

[5] Sukhpal Singh and Inderveer Chana, "Energy based Efficient Resource Scheduling: A Step Towards Green Computing." International Journal of Energy, Information & Communications, vol. 5, no. 2, pp. 35-52, 2014.

[6] Inderveer Chana and Sukhpal Singh. "Quality of Service and Service Level Agreements for Cloud Environments: Issues and Challenges." In Cloud Computing, pp. 51-72. Springer International Publishing, 2014.

Sukhpal Singh Gill *obtained the Degree of Master of Engineering in Software Engineering from Thapar University, Patiala and B.Tech. in Computer Science and Engineering. Mr. Singh received the Gold Medal in Master of Engineering in Software Engineering. Presently he is pursuing Doctoral degree in Cloud Computing from Thapar University, Patiala. Mr. Singh is on the Roll-of-honor being DST Inspire Fellow as a SRF Professional. He has done certifications in Cloud Computing Fundamentals, including Introduction to Cloud Computing and Aneka Platform (US Patented) by ManjraSoft Pty Ltd, Australia and Certification of Rational Software Architect (RSA) by IBM India. His research interests include Software Engineering, Cloud Computing, Operating System and Databases. He has more than 20 research publications in reputed journals and conferences.*