2013 IEEE Sixth International Conference on Cloud Computing

# SLA approach for "Cloud as a Service"

Ines Ayadi and Noe¨mie Simoni
Engineering School Telecom ParisTech
46, rue Barrault, 75013 Paris
Email: (ines.ayadi, simoni)@telecom-paristech.fr

Tatiana Aubonnet
CEDRIC Laboratory CNAM
292, rue Saint-Martin, 75003 Paris
Email: tatiana.aubonnet@cnam.fr

Abstract— One of slogans of cloud computing is to pay only what you use and in the highly concurrencial today context, it is important to properly identify the offer of cloud provider. The offer should not only encompass the offering services, but also the management capabilities of cloud providers. These capabilities rest on the ability to manage Elasticity, Scalability, High Availability, and On-demand Provisioning. In this paper, we focus on the description of Cloud SLA specifications by presenting cloud management capabilities "as a service" which are agreed and negotiated in the SLA contract.

## I. Introduction

Deploying Virtual applications (Vapps) in the cloud computing gives rise to several challenges mainly in terms of management that lead to higher operating costs. Many Cloud Management Constraints (CMC), specific to each cloud provider, can significantly impact the performance of the provided Vapps. These constraints are related to the ability to manage Elasticity, Scalability, High Availability, and On-demand provisioning, that should fit the Vapps' requirements. These requirements are defined by the user throughout the SLOs (Service Level Objectives) in the SLA contract. To meet these objectives, the cloud provider should not only provide customized cloud services (software, storage, computing), but also appropriate cloud management services that adjust to the Vapps needs. For this purpose, the offer of cloud provider should also include cloud management functions "as a service". Furthermore, to satisfy a perfect matching between the offer and the demand, the offering services (software, storage, management, etc.) should be described by its functional and non-functional aspects called the QoS (Quality of service). We propose in this paper a generic SLA specification for the "Cloud as a service" context. This specification supports new terms and conditions under which cloud services are managed and provided. It identifies the user's SLOs covering the E2E application's requirements. Specific management services (Pcloud services) are then defined in order to achieve these SLOs. The remainder of this paper is organized as follows. In Section II, we present the related work. Section III discusses the "Cloud as a Service" approach. In Section IV, we present the cloud SLA model. Finally, we close the paper with the conclusion.

## II. Related Work

Several research works [6], [3], [4] and commercial cloud providers (Amazon, Rackspace, Microsoft Azure, etc.) focus on the description of SLA specifications in the cloud computing. The above approaches offer a promising way to set up SLA enforcement. However, they do not reflect an accurate performance of the offered services. So, how should the Cloud provider expose its services? And must he not differentiate QoS (non-functional part) of the offered services?

Moreover, guarantees of Cloud provider's capabilities are predicted in some public clouds. For example, the availability of Amazon EC2 cloud services equals at least to 99.95%; the Amazon S3 provides a high reliability (99,99%) of storage services; Rackspace ensures 99.99% availability of load balancing services. However, other aspects are not yet addressed including scalability, elasticity, on-demand provisioning, etc. In this context, could we say that the current solutions help users to choose the adequate provider that responds to its requirements?

## III. "Cloud as a service"

1) Analysis of cloud management capabilities: In this section, we discuss some cloud management constraints that impact the performance of the deployed Vapps. Particularly, in table I, we show the factors related to cloud management capabilities including Scalability, Elasticity, High Availability and On-demand Provisioning. We conclude that cloud providers should change the way to expose its management capabilities in order to allow the choice, the customization and the wide use of them.

2) Cloud management "as a service": In order to have a well-built business policy, the "as a service" model is an adequate solution, even for management functions. So, the mentioned management capabilities (Table I) are exposed as services which can be offered, customized and negotiated in the SLA contract. In Table II, we present these capabilities a set of management services ("Pcloud services") which are specific to each cloud provider.

TABLE II
Description of Pcloud services

| Management capabilities | Pcloud Services |
|---|---|
| Elasticity | Dynamic reconfiguration and re-provisioning |
| Scalability | QOS based deployment |
| High Availability | Ubiquitous service |
| On demand provisioning | QoS-based placement and provisioning |

## IV. Generic SLA model

In the previous section, we have shown that the offerings (usage and management) of cloud provider are presented "as

978-0-7695-5028-2/13 $26.00 © 2013 IEEE
DOI 10.1109/CLOUD.2013.127

TABLE I

CLOUD PROPERTIES

| Properties | Provider factors | Existent solutions |
|---|---|---|
| Elasticity | Fastness | It depends on [1] the time required: (1) to add or remove resources (VM, LB, etc.); (2) to make a decision about the number of VM to add or release. |
| | Cost | The poor estimation (over or under provisioning) of resources number can impact the operating costs. |
| | Reactivity | To manage the load fluctuations, cloud providers systems may be [1] predictive or reactive. |
| Scalability | Initial deployment | It can be performed manually (human intervention) or dynamically (pre-defined package) [2]. |
| | Redeployment | It is performed manually that brings to application shut-down [2]. |
| High availability | Resilience | It is performed by two mechanisms: fast recovery and fault avoidance [7]. |
| | Reliability | It is performed by several ways [7]: Fault tolerance, redundancy, VM migration, etc. |
| On-demand provisioning | Automation | Making automation systems is related to: the hypervisor control and monitoring techniques and the Self-adaptation of cloud resources and services. |
| | Optimization | It aims to achieve: energy-efficient, cost-efficient, resource utilization-efficient, etc. |
| | Networking | Depends on the I/O bandwidth, the latency, the data center network architecture, etc. |

a service". These services are agreed and negotiated in the SLA contract. We define the SLA as a document which specify a personalized deal between a service provider and a consumer. It describes the terms and conditions under which cloud services are offered. In this work, we propose that the cloud provider offers many levels of SLA contract including Bronze, Silver, Gold and Platinum. This contract is created in the negotiation phase between providers and consumers. The user firstly chooses an SLA level and expresses his SLOs based on its strategies and needs (QoS requirements, price willing to pay, etc.). Once chosen, the cloud provider exposes both the Pcloud services according to the SLA contract level and the usage services according to the required QoS. Following that, the cloud consumer can select the services (usage and Pcloud) that respond to its needs.

Each SLA level encompasses a set of cloud services that have specific QoS properties. In other words, if two services (S1 and S2) have the same functionality and different behavior (QoS), we can say that $S1 = S2$. For example, if the processing delay of a service S1 takes more time than the service S2 because its algorithms are different, S1 may belong to a Silver SLA contract while S2 will be exposed in the Gold SLA contract. Consequently, the offered cloud services are classified according to its provided QoS. The non-functional part of cloud services is described by four QoS criteria: Availability (the portion of time in which a cloud service is running without failures); Reliability (the compliance rate of the deployed services); Capacity (service capabilities); and Delay (the total time to process the service). We propose in figure 1 a SLA specification for cloud computing. We are focused on the WSLA patterns [5]. New terms are added including SLACost, E2EManagementAction, Obligation, constraints (strategic, financial, juridical, etc.), etc.

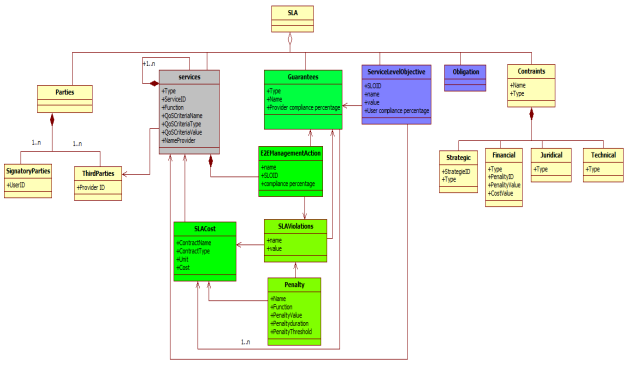


Fig. 1. SLA generic model (UML diagram)

## V. CONCLUSION

In this paper, we have presented an innovative approach "Cloud as a service" where everything is a service. This approach provides, in addition to the usage cloud services (storage, computing, software), cloud management functions as services (Pcloud services). These services are exposed and negotiated in the SLA contract. We also defined a complete SLA specification that accurately defines usage terms and conditions under which cloud services are provided.

## ACKNOWLEDGMENT

This work is supported by the OpenCloudware project. OpenCloudware is funded by the French FSN (Fonds national pour la Socie´te´ Nume´rique), and is supported by Poˆ les Mina- logic, Systematic and SCS.

## REFERENCES

[1] A. Ali-Eldin, M. Kihl, J. Tordsson, and E. Elmroth. Efficient provisioning of bursty scientific workloads on the cloud using adaptive elasticity control. In Proceedings of the 3rd workshop on Scientific Cloud Computing Date. ACM, 2012.

[2] R. Han, L. Guo, Y. Guo, and S. He. A deployment platform for dynamically scaling applications in the cloud. In Cloud Computing Technology and Science (CloudCom). IEEE, 2011.

[3] K.T. Kearney, F. Torelli, and C. Kotsokalis. Sla: An abstract syntax for service level agreements. In Grid Computing (GRID), 2010 11th IEEE/ACM International Conference on, pages 217–224. IEEE, 2010.

[4] Y. Kouki, T. Ledoux, et al. Csla: a language for improving cloud sla management. In Proceedings of the International Conference on Cloud Computing and Services Science, pages 0–0, 2012.

[5] H. Ludwig, A. Keller, A. Dan, R.P. King, and R. Franck. Web service level agreement (wsla) language specification. IBM Corporation, pages 815–824, 2003.

[6] P. Patel, A. Ranabahu, and A. Sheth. Service level agreement in cloud computing. In Cloud Workshops at OOPSLA, 2009.

[7] John Rhoton and Risto Haukioja. Cloud Computing Architected: Solution Design Handbook. Recursive Press, 2011.