A Study on Scalability of Services and Privacy Issues in Cloud Computing

R.S.M. Lakshmi Patibandla, Santhi Sri Kurra, and Nirupama Bhat Mundukur

Department of MCA, School of Computing, Vignan University,
Vadlamudi, Guntur, AndhraPradesh, India
{patibandla.lakshmi,srisanthi,nirupamakonda}@gmail.com

Abstract. Cloud Computing is rapidly emerging and the new development in Information Technology. There are many patterns, or categories, in the world of cloud computing that are needed for the enterprise architecture. Some of the categories of services are storage, database, information, process, application, platform, integration, security, privacy, management/governance, testing, and infrastructure. Scalability is one of the important features applied on any of the services. The existing analysis specially focuses on Architectural and policy Implications without exploring the data privacy issues. In this paper, the application scalability and data privacy initiatives on various services in cloud environments are presented, with an overview of the trends they follow.

Keywords: Services, Scalability, Privacy, Cloud Environment.

1 Introduction

Cloud computing is commonly associated to offering of new mechanisms for infrastructure provisioning [2,1]. The illusion of a virtually infinite computing infrastructure, the employment of advanced billing mechanisms allowing for a pay-per-use model on shared multitenant resources, the simplified programming mechanisms (platform), etc. are some of the most relevant features. Among these features/challenges, those introduced by adding scalability and automated on-demand self-service are responsible for making any particular service something more than "just an outsourced service with a prettier marketing face" [4]. As a result of its relevance, the wealth of systems dealing with "cloud application scalability" is slowly gaining weight in the available literature [3,5, 6, 7, 8, 9, 10, 11, 12, 14]. As can be observed in the previous references, automation is typically achieved by using a set of service provider-defined rules that govern how the service scales up or down to adapt to a variable load. These rules are themselves composed of a condition, which, when met, triggers some actions on the infrastructure or platform. The degree of automation, abstraction for the user (service provider) and customization of the rules governing the service vary. Some systems offer users the chance of building rather simple conditions based on fixed infrastructure/platform metrics (e.g. CPU, memory,

R. Ramanujam and S. Ramaswamy (Eds.): ICDCIT 2012, LNCS 7154, pp. 212–230, 2012.

etc.), while others employ server-level `1q a combinations of simple rules) to be included in the rules. Regarding the subsequent actions launched when the conditions are met, available efforts focus on service horizontal scaling (i.e. adding new server replicas and load balancers to distribute load among all available replicas) or vertical scaling (on-the-fly changing of the assigned resources to an already running instance, for instance, letting more physical CPU to a running virtual machine (VM)). Unfortunately, the most common operating systems do not support on-the-fly (without rebooting) changes on the available CPU or memory to support this "vertical scaling". It is, thus, necessary to extend infrastructure clouds to other kinds of underlying resources beyond servers, LBs and storage. Cloud applications should be able to request not only virtual servers at multiple points in the network, but also bandwidth-provisioned network pipes and other network resources to interconnect them (Network as a Service, NaaS) [18]. Clouds that offer simple virtual hardware infrastructure such as VMs and networks are usually denoted Infrastructure as a service Clouds (IaaS) [1, 17].

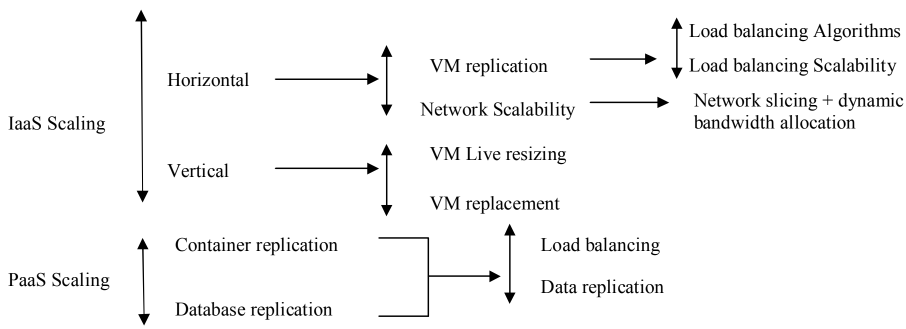

Fig. 1. Summary of the Available Mechanisms for Holistic Application Scalability

A different abstraction level is given by Platform as a Service (PaaS) clouds. PaaS clouds supply a container-like environment where users deploy their applications as software components [20]. PaaS clouds provide sets of "online libraries" and services for supporting application design, implementation and maintenance. Despite being somewhat less flexible than IaaS clouds, PaaS clouds are becoming important elements for building applications in a faster manner [1,2] and many important IT players such as Google and Microsoft have developed new PaaS clouds systems such as Google App Engine1 (GAE) and Microsoft Azure2. Due to their importance this document also discusses scalability in PaaS clouds at two different levels: container level and database level. Figure 1 provides an overview of the mechanisms handy to accomplish the goal of whole application scalability. Computing user of their private data being stolen or misused, and also assists the cloud computing provider to confirm to privacy law. Cloud computing, in which services are carried out on behalf of customers on hardware that the customers do not own or manage, is an increasingly fashionable business model. The user to the cloud, which means that they typically result in users' data, uploads the input data for cloud services being present in unencrypted form on a machine that the user does not own or control. This poses