

Chapter 3

The Challenge of Service Level Scalability for the Cloud

Luis M. Vaquero

Telefónica Investigación y Desarrollo, Spain

Juan Cáceres

Telefónica, Spain

Daniel Morán

Universidad Nacional de Educación a Distancia, Spain

ABSTRACT

This paper presents a brief overview of the available literature on distributed systems scalability that serves as a justification for presenting some of the most prominent challenges that current Cloud systems need to face in order to deliver their pledged easy-to-use scalability. Through illustrative comparisons and examples, this paper aims to make the reader's acquaintance with this long needed problem in distributed systems: user-oriented service-level scalability. Scalability issues are analyzed from the Infrastructure as a Service (IaaS) and the Platform as a Service (PaaS) point of view, as they deal with different functions and abstraction levels. Next generation Cloud provisioning models rely on advanced monitoring and automatic scaling decision capabilities to ensure quality of service (QoS), security and economic sustainability.

INTRODUCTION

Cloud computing is one of the most prominent terms employed in the latest years within the Information and Communication Technologies (ICT) field. This computing paradigm is commonly associated to new provisioning mechanisms of infrastructure resources. However, there is more

to Cloud than mere infrastructure arrangement (Vaquero et al., 2009). The Cloud aims at offering every networked resource as a service to be consumed by a variety of actors far beyond any type of administrative domain or organization.

These general ideas may be familiar to those working in any type of distributed systems technology. The Cloud was born as the harvest of many different previous seeds; it is nourished

DOI: 10.4018/978-1-4666-1879-4.ch003

by the maturity and joint of technologies such as virtualization or the Grid.

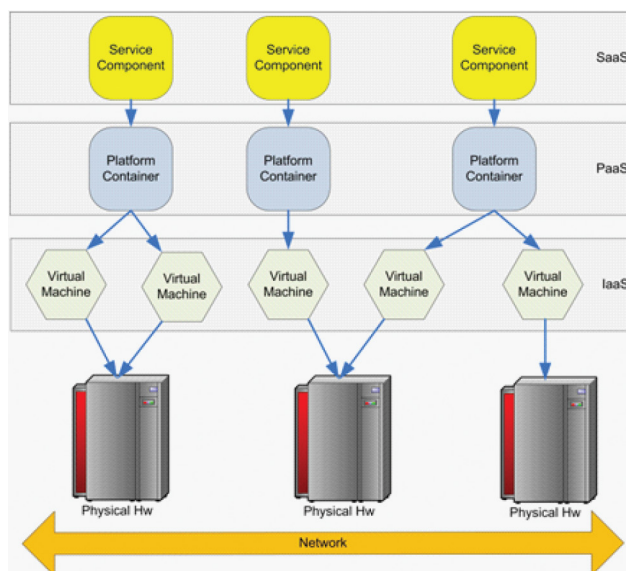
According to the functionality and abstraction level, Cloud technologies can be classified into three categories (Vaquero et al., 2009): 1) Infrastructure as a Service (IaaS), where Cloud vendors provide interfaces for the provision and management of virtualized resources (virtual machines, storage or network); 2) Platform as a Service (PaaS) which offers development tools, application containers and specialized platform technologies, and hides the details of IaaS infrastructure management; and 3) Software as a Service (SaaS), where applications are offered on demand. Currently, there are many services in the market that offer these features, for example: IaaS Clouds such as Amazon Elastic Compute Cloud (EC2) and Simple Storage Service (S3), GoGrid or Flexiscale; incipient PaaS Clouds Google App Engine, Windows Azure or Salesforce; and SaaS such as Web mail, remote desktops, Application Stores or Web 2.0 social networks.

The Cloud is already delivering important advantages that place it in an optimal position to overcome hype and evolve towards maturity.

The illusion of a virtually infinite computing infrastructure, the employment of advanced billing mechanisms allowing for a pay-per-use mode on shared multitenant resources, the easiness to create and destroy new machines, the simplified programming mechanisms, etc. are already part of our common mindset and seem to be here to stay as a new ICT utility such as the electricity or water distribution systems. Figure 1 shows how Cloud types are layered by abstraction levels and management functions. In spite of the claimed advantages, the Cloud still needs some work with regards to a variety of different subjects.

Note that an upper layer may rely on intermediate layers or may skip some of them. For example, SaaS clouds can be directly constructed on physical hardware as it has been done traditionally, of course, but they could be also deployed using IaaS functions, or PaaS over IaaS. In the purest SaaS/PaaS/IaaS Cloud model: 1) software is offered as a service and it is composed by service components which are 2) hosted in platform containers (stack of OS, middleware, application containers and libraries); platform containers run in 3) virtual machines (VMs) which are hosted in

Figure 1. The cloud stack



10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/challenge-service-level-scalability-cloud/67891

Related Content

Storage Security and Technical Challenges of Cloud Computing

Shantanu Pal (2013). *Data Intensive Storage Services for Cloud Environments* (pp. 225-240).

www.irma-international.org/chapter/storage-security-technical-challenges-cloud/77440

Green Cloud Computing with Efficient Resource Allocation Approach

Fei Cao, Michelle M. Zhuang Chase Q. Wu (2015). *Green Services Engineering, Optimization, and Modeling in the Technological Age* (pp. 116-148).

www.irma-international.org/chapter/green-cloud-computing-with-efficient-resource-allocation-approach/133061

Linked USDL Extension for Describing Business Services and Users' Requirements in a Cloud Context

Hind Benfenatki, Catarina Ferreira Da Silva, Aïcha-Nabila Benharkat, Parisa Ghodousand Zakaria Maamar (2017). *International Journal of Systems and Service-Oriented Engineering* (pp. 15-31).

www.irma-international.org/article/linked-usdl-extension-for-describing-business-services-and-users-requirements-in-a-cloud-context/191312

An Integrated Methodology for the Evaluation of Electronic Port Services

Nikolaos Marianosand Maria Lambrou (2010). *International Journal of Service Science, Management, Engineering, and Technology* (pp. 63-83).

www.irma-international.org/article/integrated-methodology-evaluation-electronic-port/45930

Measuring Interoperability Readiness in South Eastern Europe and the Mediterranean: The Interoperability Observatory

Ourania I. Markaki, Yannis Charalabidisand Dimitris Askounis (2011). *International Journal of E-Services and Mobile Applications* (pp. 73-91).

www.irma-international.org/article/measuring-interoperability-readiness-south-eastern/53535