Chapter 6
Monitoring and Metering in the Cloud

Eduardo Oliveros
Telefónica Investigación y Desarrollo, Spain

Tommaso Cucinotta
Scuola Superiore Sant'Anna, Italy

Stephen C. Phillips
University of Southampton IT Innovation Centre, UK

Xiaoyu Yang
University of Southampton IT Innovation Centre, UK

Stuart Middleton
University of Southampton IT Innovation Centre, UK

Thomas Voith
Alcatel-Lucent, Germany

ABSTRACT

Monitoring and Metering are essential activities for Service Oriented Infrastructures (SOI) and Cloud services. The information collected through monitoring is necessary to ensure the correct execution of the applications in the Cloud and the monitoring of the SLA compliance.

This chapter will present the reasons and difficulties for monitoring and metering on Cloud infrastructures. The approaches for monitoring of the execution environment and the network on virtualised infrastructures will be described together with the existing monitoring tools present on different commercial and research platforms.

INTRODUCTION

Cloud Computing Overview

Cloud computing has become a new computing paradigm as it can offer dynamic IT infrastructure, configurable service and QoS guaranteed computing environment (Wang et al., 2008). Cloud computing can be illustrated from the following aspects:

SPI model – Cloud computing originates from the concept “Hardware as a Service” (HaaS), “Software as a Service” (SaaS). Cloud now advances from SaaS to “Platform as a Service” (PaaS) and “Infrastructure as a Service” (IaaS), known as SPI model. In Cloud computing, customers can avoid capital expenditure on hardware and software by renting the usage from service provider
Monitoring and Metering in the Cloud

of third party, rather than owning the physical infrastructure by themselves. The hardware and software are rendered to customers as IT services.

**Scalability / elasticity** – Klems and Gaw (in Geelan, 2008) claim that automatic scale of infrastructure for load balancing is a key element in Cloud computing. The delivered services can elastically / dynamically grow its capacity on an as-needed basis so that the Quality of Service (QoS) can be guaranteed: “on-demand services are all cloud computing based” (de Haaff, 2008).

“Pay-per-use” / “Pay-as-you-go” / “Utility computing” – There is also a vision that Cloud computing is more like a business revolution, rather than a technology evolution. Business model, or we call “pay-per-use”, “pay-as-you-go”, and “utility computing” is another feature of Cloud computing (Geelan, 2008; Watson, Lord, Gibson et al., 2008; Buyya, Yeo, Venugopal, 2009; McFedries, 2008). The usage of the resource will be metered and service customers will pay bill to service provider for the actual resource usage.

**Data centre** - Another view of Cloud presents the data centre as the basic unit of the Cloud infrastructure (Vaquero, Rodero-Merino, Caceres and Lindner, 2009). Data centre can offer huge amount of computing power and data storage. The capacity of the data centre can change dynamically when handling a task. According to Vaquero, Rodero-Merino, Caceres and Lindner (2009), this is associated with the concept “massive data scalability” proposed by Hand (2007).

**Virtualisation** - Cloud computing can also be regarded as a “virtualised hardware and software” (Sheynkman in Geelan, 2008). This perspective emphasizes the use of virtualisation technology in the Cloud computing. Virtualization technologies multiplex hardware and have made the flexible and scalable provision of resource as hardware and software on demand easier. Virtual machine techniques, such as VMware (http://www.vmware.com) and Xen (http://www.xen.org), offer virtualized IT-infrastructures on demand. Virtual network advances, such as Virtual Private Network (VPN), support users with a customized network environment to access Cloud resources.

**Why Does Cloud Need Metering and Monitoring**

Monitoring tasks comprise a fundamental functionality in every distributed computing system. Every service should be monitored in order to check its performance and allow for corrective actions in case of failure. Monitoring data represents an operational snapshot of the system behaviour along the time axis. Such information is fundamental in determining the origin of the problems or for tuning different system components. For instance, fault detection and recovery mechanisms need a monitoring component to decide whether a particular subsystem or server should be restarted due to the information collected by the monitoring system (Litke et al., 2005).

Metering tasks are necessary for checking the disk space, network and memory usage from the machines of the platform. This information is vital to allocate services under conditions of optimum performance.

Metering and monitoring play an important role in Cloud computing, which can be attributed to the following reasons:

**From Cloud computing SPI model perspective** – Customer consumes services provided by a service provider and service provider outsources the service hosting to the dedicated infrastructure providers. Service Level Agreement (SLA) is usually employed to serve as a bilateral contract between two parties to specify the requirements, quality of service, responsibilities and obligations. SLA can contain a variety of service performance metrics with corresponding Service Level Objectives (SLO). Therefore we need to meter values of associated metrics defined in the SLA at the usage stage to monitor whether the specified service level objectives are achieved or not.

**From Cloud computing “Pay-per-use” / “Pay-as-you-go” / “Utility computing” perspec-**