Chapter 18
Reliability, Fault Tolerance, and Quality-of-Service in Cloud Computing:
Analysing Characteristics

Piyush Kumar Shukla
University Institute of Technology RGPV, India

Gaurav Singh
Motilal Nehru National Institute of Technology, India

ABSTRACT

In this chapter we are focusing on reliability, fault tolerance and quality of service in cloud computing. The flexible and scalable property of dynamically fetching and relinquishing computing resources in a cost-effective and device-independent manner with minimal management effort or service provider interaction the demand for Cloud computing paradigm has increased dramatically in last few years. Though lots of enhancement took place, cloud computing paradigm is still subject to a large number of system failures. As a result, there is an increasing concern among community regarding the reliability and availability of Cloud computing services. Dynamically provisioning of resources allows cloud computing environment to meet casually varying resource and service requirements of cloud customer applications. Quality of Service (QoS) plays an important role in the affective allocation of resources and has been widely investigated in the Cloud Computing paradigm.

1. INTRODUCTION

Cloud computing can be viewed as a model of equipping computing resources such as hardware, system software and applications as a reliable service over internet in a convenient, flexible and scalable manner. Often, this computer resources that is hardware, system software and applications are referred as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a service (IaaS), respectively. It (Buyya, Yeo, Venugopal, Broberg, & Brandic, 2009; Expósito et al., 2013) offers cost effective and effortless outsourcing of resources in dynamic service environments

DOI: 10.4018/978-1-4666-8387-7.ch018

to consumers and also facilitates the construction of service based applications equipped with the latest advancement of diverse research areas, such as Grid Computing, Service-oriented computing, business processes and virtualization.

Cloud computing providers often employs two different models to offer these services, utility computing model and Pay per Use model. Utility computing model is similar to how traditional utility services (such as water, electricity) are consumed. Whereas in Pay per Use model users are allowed to pay on the basis of number of type of service (characterized on basis of parameters like CPU cores, memory, and disk capacity)they use (A Vouk, 2008; Randles, Lamb, & Taleb-Bendiab, 2010). Payper Use model is useful in cloud resource provisioning to satisfy the SaaS user's needs with reducing cost and maximizing the profit of the SaaS provider. Another major concern for cloud resource providers is how to reduce energy consumption and thereby decreasing operating costs and maximizing the revenue of cloud provider (Berl et al., 2010; Kim, Beloglazov, & Buyya, 2009; Srikantaiah, Kansal, & Zhao, 2008).Therefore, how to serve request of the cloud services user to meet Quality of Service (QoS) needs, fault resistant reliable services and maximize the profit of the SaaS provider and cloud resource provider becomes a concern to be addressed in cloud computing environment urgently(Li, 2012).

In order to achieve its goal, Cloud, require a novel infrastructure that incorporates a high-level monitoring approach to support autonomous, on demand deployment and decommission of service instances. For this, Clouds rely greatly, on virtualization of resources to provide management combined with separation of users. Virtual appliances are employed to encapsulate a complete software system (e.g. operating system, software libraries and the deployable services themselves) prepared for execution in virtual machines (VM)(Kertész et al., 2013). Cloud management is responsible for all resources used by all the applications deployed in the cloud.

Cloud computing and networking can be viewed as the two different important keys in future Internet (FI) vision, where Internet connection of objects and federation of infrastructures become of high importance (Papagianni et al., 2013). For many cloud computing applications, network performance is a key factor in cloud computing performance to meet QoS delivery, which is directly linked to the network performance and provisioning model adopted for computational resources. Thus, in cloud paradigm the convergence between cloud computing and networking is more a requirement than a desire in order to facilitate the efficient realization of cloud computing paradigm. Providers need to consider the dynamic provisioning, configuration, reconfiguration, and optimization of both computing resources (e.g., servers) as well as networking resources to meet their objectives.

2. CLOUD COMPUTING ARCHITECTURE

Cloud computing environment supposed to furnish its huge pool of computing resources that encompasses processing power, memory, and development platform and platform to its users. This demand of sharing drives architecture of cloud computing to support convenient, efficient and flexible on demand services to users.

Architecture of cloud system comprised of different components connected in a loosely manner. These components can be broadly categorized into two parts as a front end and back end. Generally, users input and output device that includes PC, smart phone, tablet, etc. are referred as front end.

Related Content