Chapter 43
Elastic Application Container System:
Elastic Web Applications Provisioning

Sijin He
Imperial College London, UK

Li Guo
University of Central Lancashire, UK

Yike Guo
Imperial College London, UK

ABSTRACT

Cloud applications have been gaining popularity in recent years for their flexibility in resource provisioning according to Web application demands. The Elastic Application Container (EAC) system is a technology that delivers a lightweight virtual resource unit for better resource efficiency and more scalable Web applications in the Cloud. It allows multiple application providers to concurrently run their Web applications on this technology without worrying the demand change of their Web applications. This is because the EAC system constantly monitors the resource usage of all hosting Web applications and automatically reacts to the resource usage change of Web applications (i.e. it automatically handles resource provisioning of the Web applications, such as scaling of the Web applications according to the demand). In the chapter, the authors firstly describe the architecture, its components of the EAC system, in order to give a brief overview of technologies involved in the system. They then present and explain resource-provisioning algorithms and techniques used in the EAC system for demand-driven Web applications. The resource-provisioning algorithms are presented, discussed, and evaluated so as to give readers a clear picture of resource-provisioning algorithms in the EAC system. Finally, the authors compare this EAC system technology with other Cloud technologies in terms of flexibility and resource efficiency.

DOI: 10.4018/978-1-4666-6539-2.ch043

INTRODUCTION

Driven by the rapid growth of the demand for efficient and economical computational power, cloud computing (Zhang, Cheng, & Boutaba, 2010), has led the world into a new era. By enabling virtualisation technology on physical machines (PMs), it not only gives immense benefits in terms of reliability, efficiency, and scalability, but also provides virtual computational services, such as computing power, storage and network, in such a way cloud users are able to consume them over the Internet as utilities.

Most notable cloud providers, such as, Amazon EC2 (Amazon, 2010), RightScale (Adler, 2011), offer Virtual Machines (VMs) as a service to cloud users and allow the users to host their web applications on the VMs. We refer this type of approach as VM + web applications approach as shown in Figure 1(a). It allows cloud users to directly control its underlying computing resources, such as VM operations, scaling, networking, etc. In addition, this approach allows the resources of a single PM to be shared across multiple VMs for maximum efficiency. However, setting up and maintaining a working environment for web applications are complex and time consuming for cloud users, and VM resource management is a heavy-weight task for the cloud providers in this approach. In practice, we have identified two scenarios showing VM + web applications approach less feasible and less resource-efficient.

Heavyweight VM Migration: VM migration over LAN (Local Area Network) is one of the most common VM resource management operations for cloud providers. However, the VM migration over LAN is a heavyweight task. In a shared-storage environment, a VM live migration requires transferring the working state and memory from one PM to another over LAN. It consumes a large amount of I/O and network traffics in the LAN environment (He, Guo, & Guo, 2011). In a WAN (Wide Area Network) environment, mechanisms for migrating VMs remain elusive. The VM-based migration across IDCs over the Internet (Wood, Ramakrishnan, van der Merwe, & Shenoy, 2010) also requires a huge amount of I/O for both IDCs and costs a great amount of time for replicating a VM from one IDC to another. This is because a VM generally consists

Figure 1. (a) Architecture of VM + web application approach (b) Architecture of server + web application approach