

# Chapter 7

## A Computation Migration Approach to Elasticity of Cloud Computing

**Cho-Li Wang**

*The University of Hong Kong, Hong Kong*

**King Tin Lam**

*The University of Hong Kong, Hong Kong*

**Ricky Ka Kui Ma**

*The University of Hong Kong, Hong Kong*

### ABSTRACT

*Code mobility is the capability to dynamically change the bindings between code fragments and the location where they are executed. While it is not a new concept, code mobility has reentered the limelight because of its potential uses for cloud computing—a megatrend in recent years. The strongest form of mobility allows the execution state of a computational component to be captured and restored on another node where execution is seamlessly continued. Computation migration can achieve dynamic load balancing, improve data access locality, and serve as the enabling mechanism for auto-provisioning of cloud computing resources. Therefore, it is worthwhile to study the concepts behind computation migration and its performance in a multi-instance cloud platform. This chapter introduces a handful of migration techniques working at diverse granularities for use in cloud computing. In particular, this chapter highlights an innovative idea termed stack-on-demand (SOD), which enables ultra-lightweight computation migrations and delivers a flexible execution model for mobile cloud applications.*

### INTRODUCTION

Cloud computing is emerging as an important paradigm shift in how computing demands are being met in future. It is at an all-time high and actively transforming the role of IT in businesses in

recent years. Different people may have different understanding and definitions on cloud computing. But in essence, it can be seen as a model for Internet-wide access to a shared pool of hardware (processors, storage, etc) and software that are configurable on demand and presented as pay-per-use utility services, primarily *Software-as-*

DOI: 10.4018/978-1-4666-1888-6.ch007

*a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS)*. Cloud computing poses both opportunities and challenges. Opportunities center on whatever savings (in upfront expenditure, running costs, energy, etc) coming from improved resource utilization and deployment ease.

Enterprises face the problem of high management cost when running their own data centers. It is difficult to estimate the adequate scale of servers—underestimation leads to overloaded conditions while over-provisioning just means wasted investment. On-demand scalability or elasticity is an essential attribute of cloud computing that eases this trouble and reduces the total cost of ownership. With cloud architecture, computing resources can be rapidly provisioned (scale-out) and released (scale-in) with minimal management effort or service provider interaction (Mell & Grance, 2009). Such an elastic infrastructure marks the beauty of cloud computing for enhancing IT delivery's efficiency and cost-effectiveness. Challenges however include issues of data privacy, security, performance, availability and interoperability. Despite the compelling benefits of cost and flexibility, these areas need significant improvement for cloud computing to go beyond being a fledgling technology.

While enterprises, especially SMBs, are getting on the cloud bandwagon, many may still dread to become a cloud convert. Performance is increasingly becoming their paramount concern ahead of deploying their applications on clouds because it may have direct impact on quality of service and revenue. When applications move from an on-premises platform to a cloud platform, the possible increased latency could overshadow the great promise of clouds. The effect of latency on revenue has been evidenced by various key web-sites' analyses: Amazon found every 100 ms delay cost them 1% drop in sales; Google and Yahoo! got 20% and 5-9% fewer traffic from an extra load time of 500 ms and 400 ms respectively (Stefanov,

2009). The impact on latency-sensitive applications can be even more dramatic. TABB Group estimates that if a broker's e-trading platform is 5 ms behind the competition, it could lose at least 1% of its flow, analogous to a revenue loss of \$4 million per millisecond (Willy, 2008). Todd Hoff listed nine sources of latency in his article (Hoff, 2009); and among them, service dependency and geographical distribution tend to be the aggravated factors adding to extra latency when applications go hosted on clouds.

This is an expected aftereffect of adopting the cloud model which abstracts away network infrastructure details. While the abstraction can be a useful concept, it tends to weaken relationships between software components and their associated data. For example, when scaling out, or if compute cycles and storage are to be bought at their preferential rates from different cloud providers, an application may go distributed and involve more hops over the Internet for interaction, resulting in higher user-perceived latency. Therefore, any techniques that can effectively trim latency down are an important contribution to cloud computing. Collocation of computation and data access is one approach, and can be dynamically achieved by *computation migration (CM)*—an umbrella term of all the concepts behind moving an active computational component from one node to another with its execution state preserved. Besides improving data access locality, CM can also facilitate dynamic load distribution, fault resilience (migrating processes out of nodes with partial failures), and improved system administration (e.g. for server consolidation, graceful shutdown of nodes). It is a useful approach to on-demand cloud provisioning. With virtual machine (VM) migration technology, for instance, one can pool or shrink computing resources as desired by moving VM instances in or out over a cluster of nodes or even a wide area network.

The importance of mobility to cloud computing is well-established and evidenced by substantial

32 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/computation-migration-approach-elasticity-cloud/67502](http://www.igi-global.com/chapter/computation-migration-approach-elasticity-cloud/67502)

## Related Content

---

### How Hiring Baby Boomers Can Assist with the Global Cybersecurity Employee Shortage

Darrell Norman Burrell (2019). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 1-10).

[www.irma-international.org/article/how-hiring-baby-boomers-can-assist-with-the-global-cybersecurity-employee-shortage/241801](http://www.irma-international.org/article/how-hiring-baby-boomers-can-assist-with-the-global-cybersecurity-employee-shortage/241801)

### The New Profile of the Online Consumer Behaviour in a Post-Pandemic World

Vincenzo Basile, Federica Caboniand Lucia Pizzichini (2022). *Handbook of Research on Global Networking Post COVID-19* (pp. 38-54).

[www.irma-international.org/chapter/the-new-profile-of-the-online-consumer-behaviour-in-a-post-pandemic-world/309600](http://www.irma-international.org/chapter/the-new-profile-of-the-online-consumer-behaviour-in-a-post-pandemic-world/309600)

### Internet of Things Services, Applications, Issues, and Challenges

Padmalaya Nayak (2017). *The Internet of Things: Breakthroughs in Research and Practice* (pp. 491-506).

[www.irma-international.org/chapter/internet-of-things-services-applications-issues-and-challenges/177941](http://www.irma-international.org/chapter/internet-of-things-services-applications-issues-and-challenges/177941)

### Understanding Image Classification Using TensorFlow Deep Learning - Convolution Neural Network

Vinit Kumar Gunjan, Rashmi Pathakand Omveer Singh (2019). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 19-37).

[www.irma-international.org/article/understanding-image-classification-using-tensorflow-deep-learning---convolution-neural-network/241803](http://www.irma-international.org/article/understanding-image-classification-using-tensorflow-deep-learning---convolution-neural-network/241803)

### Source Coding Methods for Robust Wireless Video Streaming

Martin Fleury, Mohammad Altaf, Sandro Moiron, Nadia Qadriand Mohammed Ghanbari (2013). *Multimedia Networking and Coding* (pp. 175-207).

[www.irma-international.org/chapter/source-coding-methods-robust-wireless/73139](http://www.irma-international.org/chapter/source-coding-methods-robust-wireless/73139)