Computers and Electrical Engineering 000 (2017) 1–17

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A transparent approach to performance analysis and comparison of infrastructure as a service providers[☆,☆☆]

Bilkisu Larai Muhammad-Bello[a,1,*], Masayoshi Aritsugi[b,*]

[a] Computer Science and Electrical Engineering, Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan
[b] Big Data Science and Technology, Division of Environmental Science, Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

ARTICLE INFO

Article history:
Received 28 April 2017
Revised 15 December 2017
Accepted 15 December 2017
Available online xxx

Keywords:
Cloud performance analysis
Workload correlations
Cloud service selection
Cloud benchmarking

ABSTRACT

This study proposes a transparent approach to performance analysis and comparison of Infrastructure as a Service (IaaS) providers. Using established benchmarks, we focused on obtaining deeper insights into the performance of the public cloud infrastructure as reflected by the CPU, memory and disk I/O subsystems. We conducted experiments on a real public cloud environment and demonstrated how prospective cloud users can use our transparent methodology to discover how well virtualized public cloud resources meet their application requirements. Our transparent approach is unique in the sense that it helps prospective cloud users to decipher public cloud benchmarking data and appraise the performance of public cloud services relative to their application's performance objectives. Furthermore, we show workload correlations to the benchmarks using three real-life analytics applications. Our empirical results show significant performance differences for comparable instances on the public cloud, underscoring the need for thoughtful and transparent IaaS provider selection.

1. Introduction

The distinguishing factors between similar cloud offerings are price and performance. While the former is easy for prospective cloud users to discover and comprehend, the latter is difficult, because it requires an extensive performance analysis of the virtualized cloud resources. This challenge has motivated research in cloud service selection, where different techniques have been employed. Our previous work [1] details the research efforts undertaken thus far, elaborating on the different techniques employed in previous studies and the inadequacies identified. Cloud Service Provider (CSP) offerings are described in different ways and do not necessarily provide sufficient information regarding the true performance of the infrastructure being offered. Prospective cloud users need to be able to determine how well the public cloud instances perform relative to their on-premise systems. It is important to quantify, compare and correlate the performances of virtual

☆ This paper is an extension of the work reported in *Proceedings of the International Conference on Utility and Cloud Computing (UCC 2016) (Muhammad-Bello and Aritsugi, 2016)*.

☆☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. L. Bittencourt.

* Corresponding authors.
E-mail addresses: blabdool@dbms.cs.kumamoto-u.ac.jp, bilkisu_bello@futminna.edu.ng (B.L. Muhammad-Bello), aritsugi@cs.kumamoto-u.ac.jp (M. Aritsugi).
1 Information and Media Technology Department, Federal University of Technology Minna, P.M.B 65, Gidan Kwano, Minna Niger State, 920001, Nigeria.

https://doi.org/10.1016/j.compeleceng.2017.12.034
0045-7906/© 2017 Elsevier Ltd. All rights reserved.

compute services transparently and relative to an actual application to make an informed decision. To this end, our previous work [2] initiates a transparent framework and methodology to aid prospective cloud users in deciphering empirical benchmarking results as they relate to an on-premise application.

Different approaches have been employed in benchmarking the public cloud [3,4], and it has been argued that traditional benchmarks are inadequate in the distributed cloud context [5]. While this is true for some application-specific benchmarks, such as TPC, a number of traditional benchmarks in the area of High Performance Computing (HPC) have been found useful in cloud benchmarking [6,7,8]. However, there are still some deficiencies in the approach. Thus, we argue that the main challenges are as follows. The first is *how to make the reported benchmarking results comparable*. This is because different CSPs offer diverse services with different guarantees and capabilities. Another challenge is *how to correlate and compare reported cloud benchmark results to aid cloud users when making purchase decisions*. Thus, given the performance of a current on-premise application, we want to be able to estimate the true performance of virtual machines (VMs) on the public cloud relative to the on-premise performance. Because the information provided by the IaaS providers is unreliable, the challenge is to find a generic and informative way to estimate and describe the performance of public cloud VMs.

As one of our main contributions, we conjecture that established benchmarks are good candidates for such descriptions, and we address this challenge via the measurement methodology which is an improvement of the methodology in our prior work [2]. We observed measurements spread across the full 24 h of a day based on predefined timeslots. We also increase the frequency of observations for each experiment from 5 to 15 and take measurements both on weekdays and weekends. As a second contribution, we propose the use of observed benchmark performance scores as performance indicators and conduct performance/cost analysis and comparisons via the devised Cloud Provider Value (CPV) metric. We apply the concept of relative performance to describe how well the VMs from different IaaS providers meet the target platform VM performance. Because benchmarks can be broad in scope or application-specific, if the benchmark test used for performance evaluation is irrelevant to the target application workload, the results could be misleading or inaccurate. Therefore, as a third contribution, we address the challenge of correlating the workloads of the adapted benchmarks to real-life application workloads using three real analytics applications. Our final contribution in this paper is the study of disk I/O performance in the public cloud for optimal cloud storage selection based on the different I/O patterns.

The remainder of this paper is organized as follows. Section 2 describes the measurement methodology. Section 3 presents the empirical results and discussions. Section 4 details the performance cost analysis. Section 5 describes the workload correlations of the benchmarks to real applications. We present discussion and related work in Sections 6 and 7, respectively. Finally, Section 8 concludes the paper.

2. Measurement methodology

2.1. Subsystem components and aspects

We set out to answer the following two imperative questions. The first question relates to what is to be measured. We focus on measuring the performance of subsystem components that substantially determine an actual application's performance. Thus, we prioritized performance measurements of the CPU, memory, and disk I/O. We refrained from carrying out a comparative analysis of the performance of the disk I/O subsystems across the different providers for the reasons discussed in Section 2.6. The second question relates to how to run the experiments and take measurements. It has been found that cloud performance can be highly variable [2,9,10]. Thus, to measure this variability, we ran all tests repeatedly at different times based on the internal timing of each VM. This is because there is no accurate local time: different users will typically be in different time zones. Although we conducted all the experiments on VMs in datacenters that are physically located in the Asia/Pacific region, it was impossible to verify the locations of the users of collocated VMs. Therefore, we refrained from using external timers for our measurements. To observe the possible variability at different times of the day, we ran 5 iterations for each of the three different timeslots in a day. We defined these three timeslots to enable us to collect measurements across the 24 h of a day as follows: 3:01–11:00, 11:01–19:00, and 19:01–3:00. To observe possible variability during weekends, we repeated our experiments and ran 5 iterations during the weekend period. For each test, we ran 15 iterations on weekdays and 5 iterations on weekends.

2.2. Benchmarks details

We used synthetic benchmarks that are not application-specific, some of which are widely adopted in the field of HPC. The selected benchmark programs (described in Table 1) include the SPECjvm2008 suite,[2] Dhrystone and Whetstone from the UnixBench suite,[3] STREAM,[4] and Flexible I/O Tester (FIO).[5] The SPECjvm2008 suite comprises 10 benchmark tests provided as source code with a diverse set of workloads representing common types of integer arithmetic and floating-point computations. The operations and algorithms included in the suite are representative of generic real-world applications. We

[2] https://www.spec.org/jvm2008/.
[3] https://code.google.com/archive/p/byte-unixbench/.
[4] http://www.cs.virginia.edu/stream/ref.html.
[5] https://linux.die.net/man/1/fio.

Please cite this article as: B.L. Muhammad-Bello, M. Aritsugi, A transparent approach to performance analysis and comparison of infrastructure as a service providers, Computers and Electrical Engineering (2017), https://doi.org/10.1016/j.compeleceng.2017.12.034