# Empirical Study on the Discrepancy between Performance Testing Results from Virtual and Physical Environments

## An Extended Abstract of an Empirical Software Engineering Journal Paper

Muhammad Moiz Arif, Weiyi Shang, Emad Shihab
Concordia University
Department of Computer Science and Software Engineering
Montreal, Quebec, Canada
{mo_ari,shang,eshihab}@encs.concordia.ca

## ABSTRACT

Large software systems often undergo performance tests to ensure their capability to handle expected loads. These performance tests often consume large amounts of computing resources and time since heavy loads need to be generated. Making it worse, the ever evolving field requires frequent updates to the performance testing environment. In practice, virtual machines (VMs) are widely exploited to provide flexible and less costly environments for performance tests. However, the use of VMs may introduce confounding overhead (e.g., a higher than expected memory utilization with unstable I/O traffic) to the testing environment and lead to unrealistic performance testing results. Yet, little research has studied the impact on test results of using VMs in performance testing activities.

To evaluate the discrepancy between the performance testing results from virtual and physical environments, we perform a case study on two open source systems – namely Dell DVD Store (DS2) and CloudStore. We conduct the same performance tests in both virtual and physical environments and compare the performance testing results based on the three aspects that are typically examined for performance testing results: 1) single performance metric (e.g. CPU Time from virtual environment vs. CPU Time from physical environment), 2) the relationship among performance metrics (e.g. correlation between CPU and I/O) and 3) performance models that are built to predict system performance. Our results show that 1) A single metric from virtual and physical environments do not follow the same distribution, hence practitioners cannot simply use a scaling factor to compare the performance between environments, 2) correlations among performance metrics in virtual environments are different from those in physical environments 3) statistical models built based on the performance metrics from virtual environments are different from the models built from physical environments suggesting that practitioners cannot use the performance testing results across virtual and physical environments. In order to assist the practitioners leverage performance testing results in both environments, we investigate ways to reduce the discrepancy. We find that such discrepancy can be reduced by normalizing performance metrics based on deviance. Overall, we suggest that practitioners should not use the performance testing results from virtual environment with the simple assumption of straightforward performance overhead. Instead, practitioners should consider leveraging normalization techniques to reduce the discrepancy before examining performance testing results from virtual and physical environments.

**The full paper is published in the Empirical Software Engineering journal, and can be found at:**
https://link.springer.com/article/10.1007%2Fs10664-017-9553-x

**Please cite the following paper:** Muhammad Moiz Arif, Weiyi Shang and Emad Shihab (2017) Empirical study on the discrepancy between performance testing results from virtual and physical environments pp 1–29

## KEYWORDS

Software performance engineering, Software performance analysis, Testing on virtual environments