Experimental Support for Cloud Benchmarking and Performance Evaluation: Why Now?

A Position Paper

Sanjay P. Ahuja, Ph.D.
2010-14 FIS Distinguished Professor of Computing Sciences
School of Computing
College of Computing, Engineering and Construction
University of North Florida
Jacksonville, Florida 32224.
sahuja@unf.edu

Abstract

Today we have many public cloud providers that offer services across various cloud models such as IaaS, PaaS, and SaaS. A potential user needs to be able to objectively compare service level characteristics such as performance, scalability, elasticity, and performance variability. This is where benchmarks have a vital role to play. There is significant research activity aimed at developing new benchmarks that are suited to the cloud. This research requires intensive experimentation to evaluate the suitability of traditional benchmarks for cloud systems as well as evaluate proposed new benchmarks for the cloud. Such research requires significant cloud resources (compute, storage, and network related) depending on the application domain and generates significant empirical data. Benchmark research for big-data cloud solutions is storage intensive while HPC-type applications require significant compute resources in terms of VM instances. While public clouds can possibly be used for some of this research, in addition to the costs involved, multi-tenancy on public clouds means that workloads on the physical hosts hosting the VM instances under test can vary significantly. Dedicated private cloud facilities of meaningful capacity for this research may be too expensive for many universities. The proposed experimental facilities under the NSFCloud program can play a vital role in this regard by providing suitable experimental facilities thus providing much needed impetus to benchmarking and performance evaluation of cloud systems.

Background

Developing a successful benchmark is a non-trivial task. Given the relative newness of the cloud computing paradigm, it is no surprise that benchmarking in the cloud is still in its infancy. There are very few benchmarks that exist today that were developed specifically for the cloud. While there are many benchmarks that permit the comparison of non-cloud platforms, benchmarks for cloud environments that factor in the use of a cloud-based virtualized environment with its inherent elasticity, are few and far between. Thus, as per the state of art, there are no clear standards for end-to-end cloud benchmarking. This presents a clear research challenge as we seek to evaluate various cloud offerings.

The Standard Performance Evaluation Corporation (SPEC) is one of the most important consortiums working towards the development of new standards for cloud benchmarking. The author is a member of the SPEC – Research Group (RG), the SPEC RG Cloud Working Group, and the SPEC RG Big Data Working Group and represents the University of North Florida in these groups. In preliminary discussions within the Cloud Working Group there is consensus that there can be no one set of benchmarks that can span the incredibly diverse range of applications that seek to use the cloud as a hosting platform. Rather, cloud benchmarks will have to be developed specifically for different application domains. Initial discussions have identified the need for cloud benchmarks in the areas of High Performance Computing (HPC), Big Data or Data Intensive computing, and benchmarks that help to assess the impact of Intrusion Detection Systems on performance in cloud environments.

Summary Position Statement

In summary, the author would like to highlight three areas where experimental support from the NSFCloud program can play a vital role in the field of cloud benchmarking research and performance evaluation:

1. Executing and evaluating benchmarks suited to HPC applications require access to significant VM instances which can often execute over a time period exceeding 24-48 hours depending on the problem scale. Both the type of VM instances (such as high CPU, high memory instances) and the number of instances of each type will need to be varied to assess the performance metrics. Evaluating scalability necessitates running benchmarks of increasing workloads over a large number of instances. Furthermore, VM workload instance predictability is a desirable feature of the cloud experimental set-up.

2. Another thrust area is the evaluation of CPU+GPU cloud based clusters for domain specific HPC applications. The development of GPU based cloud offerings and the consequent inherent benefits of agility, elasticity, availability and the pay-as-use model have opened new possibilities to the HPC community in general, and biosciences research in particular by taking advantage of the parallel processing capabilities of GPUs. The NSFCloud experimental support should include support for such experimentation so that the suitability of benchmarks from non-cloud, traditional compute environments that already exist in these highly specialized HPC domains can be ascertained for cloud environments with empirical-data driven research.

3. As cloud based big data systems are maturing, the pressure to benchmark and assess the performance of these systems is increasing. For evaluating big data analytical systems, we require diverse workloads that involve, among others, financial and credit card applications, tracking feeds from social media such as Twitter and Facebook, machine learning algorithms, information extraction and data mining, and graph analysis. Such experimentation requires working with data sets that range from a 1 TB to possibly 1 PB. An experimental set-up from the NSFCloud program can play a vital role in this regard.