Future Generation Computer Systems 61 (2016) 54–65

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

New scoring formula to rank hypervisors' performance complementing with statistical analysis using DOE

P. Vijaya Vardhan Reddy *, K. Shyamala

Department of CSE, OU, India

HIGHLIGHTS

- A New scoring formula is proposed for Hypervisors' performance.
- Simulation—Created a private cloud with CloudStack Software.
- Experiments are designed with sound statistical DOE methodology.
- Hypervisors are stressed through real-time consolidated workloads.
- Results of the scoring formula are complemented with statistical analysis.

ARTICLE INFO

Article history:
Received 27 October 2013
Received in revised form
20 January 2016
Accepted 23 February 2016
Available online 7 March 2016

Keywords:
CloudStack
DOE
Hypervisor
Private cloud
SIGAR

ABSTRACT

Hypervisors enable cloud computing model to provide scalable infrastructures and on-demand access to computing resources as they support multiple operating systems to run on one physical server concurrently. This mechanism enhances utilization of physical server thus reduces server count in the data center. Hypervisors also drive the benefits of reduced IT infrastructure setup and maintenance costs along with power savings. It is interesting to know different hypervisors' performance for the consolidated application workloads. Three hypervisors ESXi, XenServer, and KVM are carefully chosen to represent three categories full virtualized, para-virtualized, and hybrid virtualized respectively for the experiment. We have created a private cloud using CloudStack. Hypervisors are deployed as hosts in the private cloud in the respective clusters. Each hypervisor is deployed with three virtual machines. Three applications web server, application server, and database servers are installed on three virtual machines. Experiments are designed using Design of Experiments (DOE) methodology. With concurrently running virtual machines, each hypervisor is stressed with the consolidated real-time workloads (web load, application load, and OLTP) and important system information is gathered using SIGAR framework. The paper proposes a new scoring formula for hypervisors' performance in the private cloud for consolidated application workloads and the accuracy of the results are complemented with sound statistical analysis using DOE.

1. Introduction

Cloud computing as a model enables dynamic access to servers, networks, applications, services, and storage and also provides an option to pay only for what has been used [1]. The major benefits of cloud computing are flexible and scalable infrastructures, reduced implementation and maintenance costs, IT department transformation (focus on innovation than maintenance and implementation) and increased availability of high-performance applications to small/medium sized businesses.

Cloud computing model promotes the availability and is composed of four deployment models. Public cloud model usually deployed over the internet and made available to all. In the community cloud model, the cloud infrastructure shared for a specific community. Private clouds are deployed behind the firewall of a company and the cloud infrastructure is operated solely for an organization. The composition of two or more cloud models forms a hybrid cloud model. With private cloud deployment model, it creates proprietary computing architecture behind a firewall with full control over infrastructure. We have created a private cloud for the experiment.

Virtualization Technology plays an important role in the success of cloud computing. The technology enables optimization of complex IT resources in a scalable way thus the delivery of services is simplified. Virtualization is a technology that combines

* Corresponding author.
E-mail addresses: vijayvardhanreddy@gmail.com (P.V.V. Reddy), kavirajvj@gmail.com (K. Shyamala).

http://dx.doi.org/10.1016/j.future.2016.02.012
0167-739X/

or divides computing resources to present many operating environments using methodologies like hardware and software partitioning, machine simulation, emulation, time-sharing, and many others [2].

Cloud computing allows customers to reduce the cost of the hardware by allowing resources on demand. The SLA (Service Level Agreement) between the providers of cloud and the customers ensures that the service will be delivered properly [3].

Hypervisor as a virtualization layer provides an infrastructural support to multiple virtual machines above it by virtualizing hardware resources such as CPU, Memory, Disk, and NIC. Hypervisors can be categorized into three models as full virtualized hypervisor, para-virtualized hypervisor and hybrid model hypervisor based on virtualization techniques that are used in their development. VMware ESXi hypervisor uses full virtualization [4] technique as every virtual machine has a virtual BIOS and an emulated PC infrastructure. All emulated hardware for the virtual machines is emulated by the ESXi kernel to give near native performance. Citrix XenServer uses para-virtualization [4] technique which involves explicitly modifying the operating system so that it is aware of being virtualized. KVM (Kernel-based Virtual Machine) is another open-source hypervisor which uses full virtualization apart from VMware ESXi and also as a kernel driver added into Linux thus effectively uses hardware-assisted virtualization hence depicts as a hybrid model.

This paper proposes a new scoring formula to rank hypervisors' performance along with statistical analysis after collecting data using DOE [5] (Design of Experiments) methodology. Design of experiments [6] is a systematic, rigorous and statistical methodology [7] for fine-tuning the experiments. We have relied on DOE than One Factor at a Time (OFAT) [8] methodology because DOE recommends the collection of data through randomization, replication and blocking with multiple factors and levels and reduces experimental errors effectively. OFAT which uses trial and error methods and information about only one factor is available and also experimental errors are not controlled. In the experiment, a private cloud is created using CloudStack [9]. Three hypervisors VMware ESXi 4.1, Citrix Systems Xen Server 6.0 and KVM (Ubuntu 12.04 Server) are deployed as hosts in the private cloud. Three applications web server, application server and the database server are installed on three guest virtual machines (VMs) of each hypervisor. Each hypervisor is stressed with consolidated application workloads for designed experiments. For applied workloads, important system information is gathered using SIGAR [10] (System Information Gatherer and Reporter) framework. The new scoring equation is proposed to rank the hypervisors' performance using gathered system information for consolidated application workloads. Scoring results are complemented with statistical analysis using DOE. Based on the score from proposed equation and with statistical analysis, we ranked hypervisors' performance in the private cloud. Three hypervisors are carefully chosen to represent full, para, and hybrid virtualization techniques. Consolidated workloads are generated using ApacheBench, JMeter, and Mysqlslap.

The paper evaluates the performances of three hypervisors. Citrix XenServer is a server virtualization platform built on the Xen Hypervisor. Xen [11] uses a para-virtualization technique. Para-virtualization modifies the guest operating system. XenServer provides a virtual infrastructure solution [12]. VMware ESXi Hypervisor uses full virtualization [13] technique. The hypervisor installs all the hardware drivers and related software into the guest operating system. It traps every instruction that attempts to update hardware data structures [14]. Hence, an extra level of mapping is in the page table [15]. KVM is a hybrid model hypervisor. KVM provides virtualization capability to guest process with user and kernel modes [16]. It uses all Linux capabilities in memory and I/O scheduling without reinventing the wheel. KVM uses hardware-assisted virtualization capabilities along with full virtualization technique.

The discussion in this paper should help both IT decision makers and end users to choose the right virtualized hypervisor for consolidated workloads in their private cloud environments based on the ranking. Statistical analysis with DOE helps in optimizing and recommending the appropriate hypervisor for server virtualization in the data center.

2. Related work

The following papers are cited to study the relevant work which had happened in the selected research area.

'A Performance Comparison of Hypervisors' [17] paper by VMware conducts different performance tests to measure the performance and scalability of two hypervisors ESX and Xen. 'A Performance Comparison of Commercial Hypervisors' [18] paper by XenSource also conducts the same performance tests to evaluate the performance of both hypervisors ESX and Xen. In the experiments, to evaluate CPU performance of two hypervisors for CPU intensive applications, they have used Standard Performance Evaluation Corporation's (SPEC) [19] SPECcpu2000. To evaluate CPU and Memory performance of hypervisors for typical system workloads they have used Passmark benchmark. They have used SPECjbb2005 to evaluate hypervisors performance for an application server's workload and SPECcpu2000 INT to assess the two hypervisors performance for development workloads. They have also used Netperf to evaluate network performance of both hypervisors. In the results, they compared both hypervisors with native and claimed that both hypervisors give near native performance except ESX scoring slightly better performance over Xen.

'Benchmark Overview—vServCon' paper by Fujitsu PRIMERGY Servers [20] talks about 'vServCon' benchmark which was developed for their internal purpose to measure and assess the performance of virtualized servers. According to them, vServCon is not a new benchmark, but a framework that consolidates already established benchmarks, as workloads, in order to simulate the load of a virtualized consolidated server environment. Three applications database, application server, web server executed in each virtual machine and all these applications are stressed with load generators through established benchmarks. All individual results are summarized into one result and they named it as a score for the performance capability of a virtualized environment.

Different hypervisors such as XEN, KVM, and VMware ESX [21–23] performances have been evaluated to measure the overhead of virtualization using multiple toolkits. Menon had used a toolkit called Xenoprof (system-wide statistical profiling toolkit for Xen virtual machine environment similar lines of OProfile for Linux) to evaluate the performance overhead of network I/O devices. Menon [22] had used Xenoprof to debug Xen and been able to improve the network performance. Menon claims from his research that domain0 performance is close to native, but guest operating system performance degrades considerably because of high CPU utilization as virtualization increases the number of instructions that are to be handled by CPU. Jianhua [23] had used LINPACK benchmark tool to test processing efficiency on floating point. Jianhua had observed that windows XP gives better performance than fedora 8 on Xen. Jianhua had clarified it as Xen owns certain enhancement packages for windows XP than fedora. Jianhua had used LMbench to evaluate memory virtualization of Xen and KVM and noticed that Xen's performance is better than that of KVM. Jianhua had used IOzone to compare file system performance among Xen and KVM. Jianhua had found that without intel-VT processor the performance of Xen and KVM is