A Comparative Study on I/O Performance between Compute and Storage Optimized Instances of Amazon EC2

Abu Awal Md Shoeb, Ragib Hasan, Md. Haque, and Meng Hu
Department of Computer and Information Sciences
University of Alabama at Birmingham
Birmingham, AL 35294, USA
Email: {shoeb, ragib, mhaque, mhu}@cis.uab.edu

***Abstract*—Cloud computing infrastructure helps users to minimize cost by outsourcing data and computation on-demand. Due to the varying user needs in terms of computation power, storage capacity, etc., cloud providers offer various machines to choose from, to maximize the intended need. In this paper, we disprove several common conceptions regarding the performance and cost of cloud by experimenting on instances of two different families (compute and storage optimized) of the most popular cloud platform, Amazon Elastic Compute Cloud (EC2). Our analysis shows the interesting finding that, for the machines of the same configuration, storage optimized instances have lower disk read-write speed than compute optimized, which does not completely reflect the claim made by Amazon in all cases. Additionally, storage optimized instances have notable performance difference among them. We also identify that the I/O performance of same instance type varies over different time periods.**

***Keywords*-Cloud Computing; Amazon EC2 Performance**

I. Introduction

The use of a cloud computing system greatly reduces the cost of large infrastructure setup for users. The providers use the Infrastructure-as-a-service (IaaS) model of cloud computing, where customers rent virtual machines in the cloud and run their applications on them. In IaaS, a common problem is the lack of information and performance comparison/variation of different types of virtual machines. Performance of a virtual machine might vary with time, type of operation, file size, and file type, and so on. In order to illustrate this, we performed a case study in the Amazon Elastic Compute Cloud service (EC2). They provide a wide range of instance types with different size of memory, storage, LAN and CPU speed[1]. To make it easier for users to select the best option for their applications, Amazon EC2 instance types are grouped together into families based on target application profiles. Among several instance families, we analyze the performance of instances taken from storage optimized and compute optimized families.

According to the Amazon EC2 website, it is recommended for users to measure the performance of applications to identify appropriate instance types and validate application architecture. To select an instance, a user should not completely rely on the description of an instance provided by Amazon. Suppose, a user requires high speed I/O operations regardless of high computation power or storage capacity. An application or web service which always writes or updates customer data would be a practical example for this case. In that circumstance, the user might go for storage optimized instance as it ensures the highest I/O operations than any other instances Amazon has. However, in almost 50% cases, our experiment shows that, compute optimized instances could perform faster I/O operations than storage optimized instances. As a result, the user will pay more but might not achieve the highest performance[2].

We evaluate the performance of an instance using standard benchmark tool to compare the performance of different instances from the same type (storage or compute optimized) and determine performance variations. We also compared instances from the same type but launched at different times to determine the effect of time on performance. Finally, we compared the I/O performance of storage optimized instances versus compute optimized instances to determine whether the recommended EC2 configurations for storage-intensive applications actually perform as advertised. While researchers have performed different type of analyses involving EC2 instances, most of the studies have been done to compare instances of the same type only. To the best of our knowledge, this type of comparison is the first of its kind, which compares instances of different types to compare their performance in specific application scenarios.

II. Background

A. Amazon Elastic Compute Cloud (EC2)

Based on the price, Amazon EC2 has four types of instances. They are on demand, and light, medium, and heavy utilization. On demand instance has no upfront price where others do. Based on these pricing factors, we choose on demand instances for our experiment. Among many instance families, compute and storage optimized instances are available for on demand basis. Compute optimized instances are designed for applications that benefit from high compute power. It has a higher ratio of virtual CPUs (vCPUs) to memory than other families, and the lowest cost per vCPU among all Amazon EC2 instance types. On the other hand, storage optimized instances are designed for very high storage density, low storage cost, and high sequential and random I/O performance.

TABLE I: I/O Performance Percentage

Operation	Type	c3.xlarge	i2.xlarge
Read	Random	46%	54%
	Sequential	40%	60%
Write	Random	50%	50%
	Sequential	51%	49%
Delete	-	92%	8%

B. Benchmark Tools

The benchmarks are simple execution files and do not need any installation or setup. Results are displayed when the tests are running and performance results are saved in a text file. It measures disk write and read speeds of multiple files at different block sizes. Also it does the same for bus/DMA speeds, by repetitively reading data from the disk's buffer, and measures random reading time of various sized files.

III. Experimental Setup

We modified standard benchmark tool according to our requirements and applied to EC2. We also performed benchmark to our local test bed to make sure the functionality of our modified tool. We pick five i2.xlarge storage optimized instances and five c3.xlarge compute optimized instances from Amazon EC2. Since we consider comparing the performance between two different types of instance families, we pick instances which have similar vCPU, ECU (EC2 Compute Unit), and LAN speed. However, they have different memory size and storage. We choose US East Virginia location as it is less expensive than other regions. We run our tools in different time periods over a month. We performed benchmark for 20 times for each instance that includes read, write, and delete operations. We used both single and multiple files of 8MB, 16MB and 32MB for I/O operations. We also performed I/O operations with other file sizes ranging from 2MB to 128MB.

IV. Results

Table I demonstrates the disk performance ratio of compute optimized and storage optimized instances. This percentage is generated from the total number of operations performed on all instances. We performed around 1,200 I/O operations among all instances to determine the performance variations between these two instance families. The numbers reflect how many times an instance gives better performance than the other. It is noteworthy that almost half of the I/O operations were slower in storage optimized than compute optimized instances. Moreover, c3.xlarge seems to be dominant only for write and delete operations. However, compute optimized instance has better cache performance in every case compared to storage optimized, and hence has been ignored in our results.

Figure 1 shows performance variation of some storage optimized instances. The instances were created in different times. The read operation was done for single and multiple files of 32 MB. The figure clearly represents the difference of performance among four instances. However, it is also obvious that the same instance has almost same read speed in every case, irrespective of the file size, and single or multiple files.

Figure 2 shows the comparison of random write operation between compute optimized and storage optimized instance.

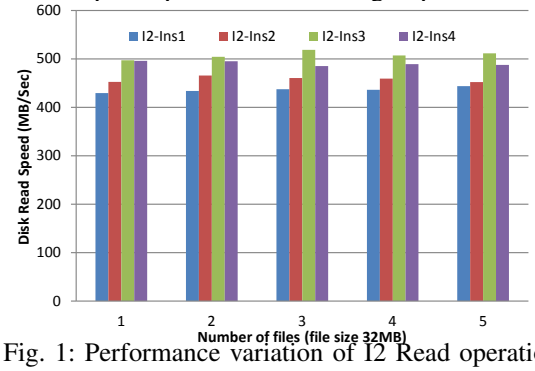

Fig. 1: Performance variation of I2 Read operation

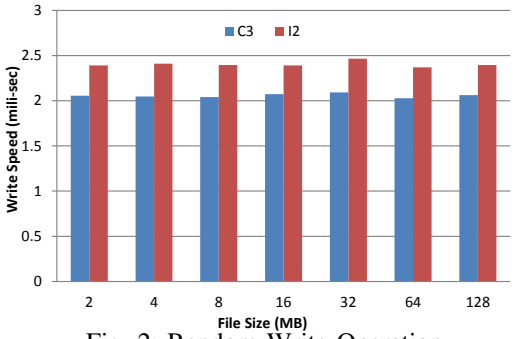

Fig. 2: Random Write Operation

1 KB blocks are read randomly from 7 file sizes between 2 MB and 128 MB. Results reflect the disk's buffer size and rotation speed. The same file size was used for write operation. For read operations, compute optimized instances took longer time than storage optimized. However, for write operation, compute optimized always took lesser time than storage optimized instances.

V. Conclusion and Future Work

This paper presents an analysis of the I/O performance differences between compute optimized and storage optimized instances of Amazon EC2. We used standard PC benchmarks to obtain measurements and describe how a compute optimized instances behaves almost same as storage optimized instances in terms of disk I/O performance. Our analysis can be utilized by cloud providers to ensure performance consistency of instances, by users to get prior idea when selecting an instance, and by researchers to optimize performance of multi-tenant systems. Our ongoing work focuses on Amazon EC2 instances running different operating systems, as well as I/O performance across different regions. In future, we plan to develop a cost model by including all features of the user's task and present the most cost effective option for cloud services.

VI. Acknowledgment

This research was supported by a Google Faculty Research Award and an Amazon AWS in Education Teaching Award.

References

[1] "Aws - amazon ec2 instance types," http://aws.amazon.com/ec2/instance-types.

[2] Z. Ou, H. Zhuang, A. Lukyanenko, J. K. Nurminen, P. Hui, V. Mazalov, and A. Yl?-J??ski, "Is the same instance type created equal? exploiting heterogeneity of public clouds," *IEEE Transactions on Cloud Computing*, vol. 99, no. PrePrints, p. 1, 2013.