Relative Cost-to-Performance Analysis of Virtual Machine Types in Public Cloud Systems

Jun-Hyeong Choi 1 and Jong Wook Kwak 2

1Research Scholar, 2Associate Professor,
1,2 Department of Computer Engineering, Yeungnam University,
280 Daehak-Ro, Gyeongbuk 38541, Republic of Korea.

1Orcid ID: 0000-0001-8932-7162, 2Orcid ID: 0000-0002-6639-3738

Abstract
Recently, as the role of users has expanded from service consumers to information providers, a need for a system that processes a lot of information has arisen. Small and medium-sized service providers have a difficulty building systems by purchasing equipment due to initial installation costs and maintenance costs of the system. Therefore, small and medium sized service providers build systems using public cloud systems. In this paper, we conducted experiments on the CPU and disk aspects to evaluate the performance of the public cloud system. The CPU is the most influential factor in system performance, and the processing speed of the data depends on the performance of the CPU. In addition, the public cloud system is used not only for constructing the system that processes operations but also for storing a lot of data, so performance evaluation of the disk is also necessary. In this paper, we have created virtual machines by defining performance evaluation modeling to measure the performance of the public cloud system. Performance of CPU and disk was measured using benchmark. Using the CPU performance measurement results, we compared the price and performance gains of the public cloud system provider. Mostly, service providers showed the highest performance gains in mid-sized system. In the environment of CPU operation and application execution, low and mid of GCP are more advantageous than AWS, but AWS is more advantageous at high. Although AWS performed well when using virtual machine for data storage, it should be considered that the price of AWS is 152% higher than GCP on average.

Keywords: Public Cloud System, Cloud Computing, Virtual Machine, CPU Utilization, Disk Read/Write Time

INTRODUCTION
Recently, the development of web services has enabled users to use various types of services easily and conveniently. Particularly, with the development of SNS, users have expanded their roles as providers of simple information. The expansion of user roles has had a major impact on Web services and developers. The web service provider plays a role of collecting, analyzing and reprocessing the information received from the user while providing good quality service to the user. High-performance machines are required to collect and analyze large amounts of information. However, there are a lot of initial costs to build a system using high-performance machines. This makes it difficult for individual or small-scale web service providers to build high-quality systems, so they use public cloud systems.

The public cloud system has the characteristics that the web service provider rents computer resources through the web without purchasing the components of the system directly. Resources that can be borrowed from a public cloud system include processors, main memory, and disks. Web service providers can use the characteristics of public cloud systems to save capital used in building and maintaining systems that provide web services. It is possible to provide high-quality service using a low cast.

Public cloud services can be roughly divided into three types according to the model in which the service is provided. IaaS, PaaS, and SaaS. IaaS (Infrastructure as a Service) provides services to use machines, repositories, databases, and networks to be used as servers. Platform as a Service (PaaS) provides a platform for development, and Software as a Service (SaaS) is a service that provides the necessary software through the Web. As mentioned earlier, there are
various types of services in cloud computing. Typical commercial services are Google's Google Compute Engine and Amazon's Amazon Web Service.

In this paper, we evaluated the performance of IaaS, the most common usage of public cloud systems. We have created virtual machines by defining performance evaluation modeling to measure the performance of the public cloud system. Performance of CPU and disk was measured using parsec benchmark.

Using the CPU performance measurement results, we compared the price and performance gains of the public cloud system provider. Both providers showed the highest performance gains in medium instance. In the environment of CPU operation and application execution, small and medium instance of GCP are more advantageous than AWS, but AWS is more advantageous at large instance. Although AWS performed well when using virtual machine for data storage, it should be considered that the price of AWS is 152% higher than GCP on average.

BACKGROUND

Cloud Computing

The cloud computing means borrowing expensive hardware or software from the data center over the internet at a low price. Individuals or developers who want to provide web services lease their own software or virtualized hardware resources through the data center. At this time, the user pays the price of the resource used by the user, which is inexpensive as compared with actually purchasing and operating the high-performance resource directly [1][2][3].

Using cloud computing, users can freely use the systems they want to build within the resources leased from the data center. Therefore, in the characteristics of advanced cloud computing, users can utilize various forms from simple document work to construction and operation of web server.

The Classification of Cloud Computing

The cloud computing can be divided into 3 categories according to the level of resource leasing, and the range that users can control by each service is shown in figure 1 [4][5].

First, IaaS(Infrastructure as a Service) refers to a service that borrows computer resources from a virtualized data center such as a server or a repository, or builds an infrastructure using the Internet. In such a case, the user selects a machine according to the type of service to be provided. After that, it is possible to build and use the infrastructure of the form necessary for the user like the storage and the network. In this case, the user can directly control most resources required for building the server.

Second, PaaS(Platform as a Service) is a service for software developers and provides a platform for developers to develop software. It is mainly used when there is no environment to save the development tools or when testing the development results.

Finally, SaaS(Software as a Service) is a service that provides the user with the environment for using specific software on the Web. A service in SaaS is that provides the environment of the software on the web without installing the software on the user's PC. For example, when you want to work on a document, you may not have a word processor program installed on your PC. In such a case, the user is provided with an environment in which the user can access the software through the web site.

RELATED WORKS

The public cloud system can be configured in various forms according to the user's purpose. When a user collects, analyzes, and processes data using a public cloud system, a system with a high computing power is required.

In addition, a system that needs to accommodate a large number of clients requires a public cloud system that has excellent computing power and excellent network performance. Therefore, in this section, we examine the comparative analysis of public cloud systems from various perspectives.

Nawaz et al. compares and analyzes the performance of cloud storage through astronomical data, which has a large size and generates intermediate files during processing. In this study, the types of cloud storage are classified into Cloud storage, VM storage and Submit host. Experiments were performed on the transfer time of intermediate files and the execution time of workflow, but the performance evaluation of disk and memory was insufficient [6][7].

Persico et al. compared network performance between AWS and Azure's data center in a public cloud system. They measured TCP throughput, UDP throughput, and RTT according to the location of the data center. As a result, They analyze and evaluate network performance of AWS and Azure [8].

Unlike the previous two studies, Cheng et al. measured the performance of VMs in public cloud systems. In addition, we compare the price and performance measurement results of public cloud system providers and propose a correlation between price and performance of the public cloud system. Public cloud system providers were divided into two groups according to their size, and performance was measured in terms of CPU, disk, and memory. As a result, small-scale public cloud system providers are more stable and perform better than large-scale providers [9].
Challenges of Public Cloud Systems

As mentioned earlier, public cloud systems have a variety of features and advantages that are convenient to use, but there are still many problems to overcome.

According to researches, there are five things that need to be improved in public cloud systems today. It should be easy to manage and utilize the infrastructure and provide services at low prices. In addition, there should be no limit to the use of the service. The management of sudden accidents is needed, and the waste generated from the use of the system should be reduced [10][11].

Besides, a security is an important factor for the widespread use of public cloud systems. The public cloud system shares the information because of its characteristics, and this information includes the personal information of the individual. Therefore, a security policy with high reliability is required for the generality of the public cloud system [12][13].

METHODOLOGY

Performance Comparison of Public Cloud Systems

The public cloud system is provided in the form of IaaS, PaaS, SaaS according to the user's convenience. In addition, when constructing a system, it does not purchase hardware and software components directly, but rents the necessary amount to pay instead. At present, public cloud systems are provided by many companies, including Amazon's AWS, and Google's GCP.

Each public cloud system provider provides different types of virtual machines to suit their purposes. It can be classified into standard type, high performance CPU, and large memory type, and the performance of the virtual machine can be selected according to the user's demand within the category. Table 1 shows the standard virtual machines provided by each public cloud system divided into small instance, medium instance, and large instance according to the performance and the prices, accordingly. The prices in Table 1 are based on July 2017 for virtual machines using the Linux operating system. Although each cloud system vendor offers different types of virtual machines, it can be seen that the price of Google's GCP is generally low. Amazon's Amazon Web Service, for example, has a relatively higher price compared to Google's GCP because it includes its own storage system in the cost of providing virtual machines. Therefore, it can be seen from Table 1 that when performing only high-performance computation without service such as automatic backup of data, it can be seen that Google's GCP is more advantageous in terms of price than Amazon's Amazon Web Service.

Table 1: The type of virtual machines

<table>
<thead>
<tr>
<th>Service Provider</th>
<th>Instance Type</th>
<th>Machine Type</th>
<th>CPU</th>
<th>Memory Size(GB)</th>
<th>Price / Month(USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Cloud Platform</td>
<td>Small Instance</td>
<td>n1-standard-1</td>
<td>2.5GHz, 1 core</td>
<td>3.75</td>
<td>33.77</td>
</tr>
<tr>
<td></td>
<td>Medium Instance</td>
<td>n1-standard-2</td>
<td>2.5GHz, 2 cores</td>
<td>7.5</td>
<td>67.54</td>
</tr>
<tr>
<td></td>
<td>Large Instance</td>
<td>n1-standard-4</td>
<td>2.5GHz, 4 cores</td>
<td>15</td>
<td>135.08</td>
</tr>
<tr>
<td>Amazon Web Service</td>
<td>Small Instance</td>
<td>t2.small</td>
<td>2.4GHz, 1 core</td>
<td>2</td>
<td>41.87</td>
</tr>
<tr>
<td></td>
<td>Medium Instance</td>
<td>t2.large</td>
<td>2.4GHz, 2 cores</td>
<td>8</td>
<td>116.14</td>
</tr>
<tr>
<td></td>
<td>Large Instance</td>
<td>t2.xlarge</td>
<td>2.4GHz, 4 cores</td>
<td>16</td>
<td>217.84</td>
</tr>
</tbody>
</table>

To analyze and evaluate the performance of each virtual machine, we used Parsec Benchmark Suits 2.1. Parsec Benchmark Suits 2.1 includes various types of workloads. Depending on the characteristics of each workload, the performance of a virtual machine can be interpreted in various forms.

Figure 2: The structure of the proposed model for performance measurement

In this paper, we implemented a VM stat, the VM status measurement module that records the state of the virtual machine in addition to executing the workload included in Parsec Benchmark Suits 2.1 in the virtual machine. The
The operation procedure of the VM stat Measurement Module is shown in Figure 2. Running the VM stat Measurement Module will fill all areas of the main memory for once. This prevents excessive performance enhancement of the main memory due to the large amount of idle space. Also, if the data for the corresponding workload is already stored in the main memory, it will affect the fair performance evaluation. After the above operation is completed, the Workload Run Thread in Figure 2 executes all the workloads in the virtual machine sequentially. Status Measurement Threads record the state of the CPU, memory, and disk of the virtual machine while the Workload Run Thread executes the workload in the virtual machine. When one workload is completed, the Status Measurement Threads stop recording the status of the virtual machine and save all the data recorded so far to a file. This process is repeated until all the workloads are executed. When all the workloads are completed their execution, the stored files are collected.

![Figure 3: The normalized CPU utilization](image1)

![Figure 4: The normalized runtime](image2)
Performance Evaluation

Experiment Environment

In this paper, we compare and analyze the performance of public cloud systems using PARSEC benchmark suits 2.1 [14]. PARSEC (Princeton Application Repository for Shared-Memory Computers) benchmark suits 2.1 is a benchmark produced by Intel and Princeton University. PARSEC benchmark suits 2.1 includes applications in various fields such as finance, computer vision, and multimedia, helping to identify the public cloud system that is suitable for the user's purpose. The input set of the PARSEC benchmark suits 2.1 is divided into test, simdev, simsmall, simmedium, simlarge, and native depending on the size of the problem. In this paper, the experiment was performed by selecting the input set which is the largest of six input sets and reflecting the actual usage environment. As mentioned earlier, the PARSEC benchmark suits 2.1 includes various types of workloads such as finance, computer vision, and multimedia. Table 3 shows the magnitude of the problem for each workload [15].

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Problem size in native</th>
</tr>
</thead>
<tbody>
<tr>
<td>blackscholes</td>
<td>10,000,000 options</td>
</tr>
<tr>
<td>bodytrack</td>
<td>4 cameras, 261 frames, 4,000 particles, 5 annealing layers</td>
</tr>
<tr>
<td>canneal</td>
<td>15,000 swaps per temperature step, 2,000° start temperature, 2,500,000 netlist elements</td>
</tr>
<tr>
<td>facesim</td>
<td>80,598 particles, 372,126 tetrahedra, 100 frames</td>
</tr>
<tr>
<td>ferret</td>
<td>3,500 image queries, database with 59,695 images, find top 50 images</td>
</tr>
<tr>
<td>fluidanimate</td>
<td>500,000 particles, 500 frames</td>
</tr>
<tr>
<td>freqmine</td>
<td>Database composed of spidered collection of 250,000 web HTML documents, minimum support 11,000</td>
</tr>
<tr>
<td>streamcluster</td>
<td>1,000,000 imput point, block size 200,000 points, 128 point dimensions, 10-20 centers, up to 1,000 intermediate centers allowed</td>
</tr>
<tr>
<td>swaptions</td>
<td>18,000 × 18,000 pixels</td>
</tr>
</tbody>
</table>

Figure 5: The normalized total time of read operation
**CPU Performance Analysis**

Figure 3 shows CPU utilization during the workload execution in order to measure CPU performance. As shown in figure 3, the runtime is longer as CPU utilization is lower.

Figure 4 (a) shows runtime of AWS compared with GCP. As a result, the average runtime of AWS is 573.6%, 141.2% and 99.44% according to all specifications. Low and mid of AWS were low performance due to low hardware specification compared to GCP, but it has 0.56% performance increase at high due to large memory size compared to GCP high.

Figure 4 (b) shows the runtime performance of the GCP in the low, mid and high virtual machines based on the GCP low. As the instance type of virtual machine enhances in mid and high compared to low, the average performance increase of mid and high is 8.61% and 10.68%, respectively compared to low.

Figure 4 (c) shows the runtime performance of AWS. Like the results of figure 4(b) virtual machines of mid and high enhances the performance about 79.36% and 86.18% respectively, compared to AWS-low.

Both vendors showed the highest performance at high instance, among the specifications. However, relative performance increase of virtual machine between low and mid is more than that of one between mid and high. Besides, as the performance of instance type increases, the maximum performance of AWS is higher about 152%, compared to GCP high.

**Disk Performance Analysis**

Figure 5 (a) shows the total read operation time of GCP and AWS, and it was 27.41%, 102.03%, and 71.47% in AWS according to the specifications based on the total time required for GCP read operation, respectively. Figure 5 (b) shows the result of comparing the total read operation time in mid and high virtual machine of GCP based on GCP low. As a result...
of comparing the total read operation time of mid and high based on GCP low, the total read operation time was 20.15% and 23.54%, respectively.

Figure 5 (c) shows the result of comparing the total read operation time on the AWS mid and high virtual machine based on AWS low. As a result of comparing the total read operation time of mid and high based on AWS low, the total read operation time was 75.01% and 61.38%, respectively. Both vendors showed the lowest performance at low, which are attributed to frequent read operations due to small memory size. Besides, the read performance of GCP showed 3.39% performance decrease from mid to high. It seems to the reason of the configuration and location of the public cloud system.

Figure 6 (a) shows the total write operation time of GCP and AWS. As a result, it was 18.1%, 18.47%, and 16.84% in AWS according to the specification based on the total time required for each GCP write operation.

Figure 6 (b) shows the result of comparing the total write operation time of mid and high based on GCP low. As a result of comparing the total write operation time of mid and high based on the GCP low, the average write operation time is 92.88% and 98.73%, respectively.

Figure 6 (c) shows the result of comparing the total write operation time of mid and high based on AWS low. As a result of comparing the total write operation time of mid and high based on AWS low, the write operation time is 94.78% and 91.87%, respectively.

Both vendors showed the lowest performance at low, but the relative performance increase between mid and high was ranged from 1.36% at minimum to 8.13% at maximum.

CONCLUSION

In this paper, we evaluate the performance of the public cloud system in terms of CPU and disk. The CPU is the most important performance indicator of the system. Public cloud systems are also used to build systems that process operations, but they are also used to store large amounts of data, so disk performance evaluation is also required. In this paper, performance evaluation modeling is defined to measure the performance of the public cloud system to create a virtual machine.

Using benchmarked CPU performance measurement results, mid and low prices of both providers showed the highest performance gains compared to the prices and performance gains offered by public cloud system providers. As a result of comparing runtime after executing workload on each provider's virtual machine, AWS showed average 573.6%, 141.2%, and 99.44% performance according to all specifications compared to GCP, and GCP generally showed a favorable aspect compared to AWS. As a result of comparing the total read operation time during the execution of the workload in each provider's virtual machine, the average results of each instance type of AWS was 27.41%, 102.03% and 71.47%, compared to each corresponding instances of GCP. As a result of comparison of total write operation time, AWS showed an average of 18.1%, 18.47% and 16.84% compared to GCP, and AWS showed a favorable aspect when using virtual machine for data storage and archiving. However, it should be considered that the price of AWS is 152% higher than GCP.

ACKNOWLEDGEMENT

This work was funded by the BK21+ program of the National Research Foundation (NRF) of Korea. (Corresponding author: Jong Wook Kwak)

REFERENCES


