Understanding the Cost of Cloud Computing and Storage

Iman Sadooghi*, Dongfang Zhao*, Tonglin Li*, Ioan Raicu*†
isadoogh@iit.edu, dzhao8@hawk.iit.edu, tli13@hawk.iit.edu, iraicu@cs.iit.edu
*Department of Computer Science, Illinois Institute of Technology
†Mathematics and Computer Science Division, Argonne National Laboratory

ABSTRACT

Cloud computing is a new paradigm that enables everyone to use the remote resources in any demanded scale. Users are able to use computing and storage resources on-demand. Big companies as well as many startups have provided large scale cloud resources to different level of users. The users of cloud resources vary from huge tech and non-tech companies to the small start-ups and individual users. Unfortunately the pricing of these products have not yet adapted to the budget limit all of the potential users. There is a big need for lower cost storage system that can be used by the individual users. Also in the computing part, many small businesses have the potential to use these resources as their building blocks and avoid infrastructure costs.

In the storage, the gap between the raw storage price and the price that is offered by the current online storage providers determines that there is a big opportunity to provide a cloud storage system that focuses on normal end-users instead of big companies.

The compute part of the cloud also has potential to grow. Looking at the offered price by the cloud service providers brings the question in mind that whether there is opportunity to provide low cost compute resources in this area.

INTRODUCTION

Based on current trends, there is a fast increase of online data in many companies and organizations [1]. Online storage is used for different usages. Using cloud storage provides a highly available data that can be accessed from different locations [2]. In science, different applications generate datasets that are growing exponentially. There is a big need for data archival for science data with a better performance than the performance of current file systems. All of these reasons show that there is a big need for online storage [3]. Also Computing Cloud has gained attraction to transform a large part of the IT industry over the internet. Companies tend to use Cloud resources instead of attempting to buy and maintain their own infrastructure. One of the important features of the cloud that makes it very interesting for users is the low cost. Cloud computing is one of the areas which has the potential to transform a large part of the IT industry over the internet.

In this work, we are going to analyze the cost of the cloud computing and cloud storage. We are interested to compare the cost of porting the computing and storage to the cloud with the traditional options. The commercial cloud providers don't reveal their information about the cost of their service. Therefore we need to come up with an accurate detailed estimation of a private cloud that includes all features of a commercial cloud. A cloud system may need to utilize many technologies to fulfill the user requirements. Security, reliability, availability and performance are some important features that a cloud system should have [4]. We make sure to include these features in our estimations. For example we estimate the storage systems with 3X data replication to have a fair comparison with the cloud storage systems.

METHODOLOGY

We include both compute and storage Cloud in our comparisons. For compute cloud, we first propose an accurate estimation of a compute intensive node that could be used for a datacenter and try to compare the cost of estimated cloud with Amazon EC2 instances. We consider the cost of power, cooling and system administration in the total cost. Likewise we propose a storage intensive node for our storage comparisons and compare it with public cloud.

Cost Comparison

Figure 1 shows the price comparison of all of the instances on Amazon EC2 and the prices of similar instance from the private cloud with the same features.

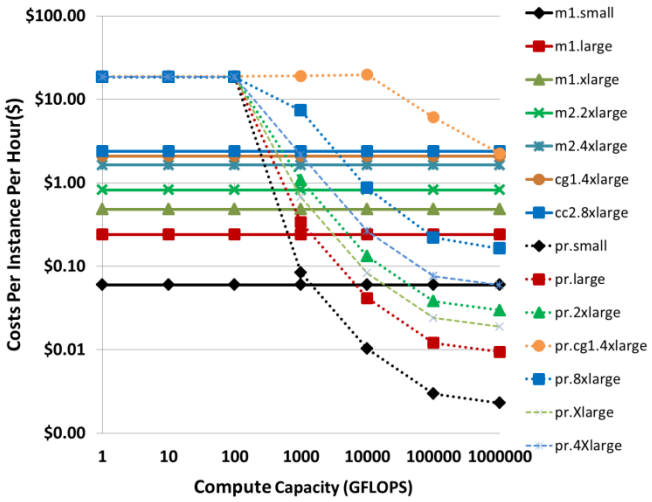

Fig. 1. Comparison of EC2 instances vs. private compute cloud estimations.

We can observe that the Amazon instances are more cost effective on smaller scales. The private cloud instances can only beat the Amazon prices on larger sizes when their total compute capacity is more than 1000 GFLOPS. The main reason for that is having some constant cost up to a certain scale. Therefore these costs break down when the cluster size is large. We can conclude that having your own private cloud is not cost effective unless the cluster size is big enough to break down the cost of system administration on each node and the utilization level of the private is higher than a certain level. Otherwise it is wiser to use the cloud instances.

Figure 2 compares the cost of Cloud storage in different providers versus having a private internal cloud storage . It compares the price of having the data in three different sizes of 100GB, 1TB and 3TB. The HDD price is the price of single hard drive. It does not include the power and other prices. As it is evident, the cost of having the raw storage is almost 100X cheaper than the costs cloud storage providers service.

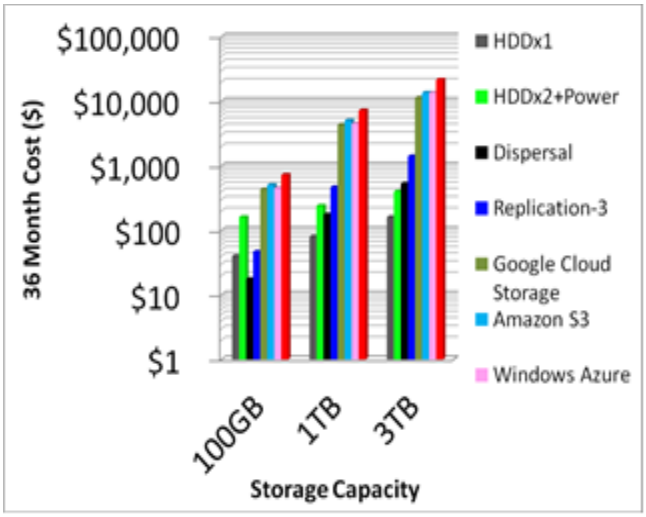

Fig. 2. Total price comparison of cloud storage different providers.

Figure 3 shows the price of different products for 1 GB in a month. The estimated dispersed storage provides the most cost effective price comparing to other systems. As the storage size gets bigger the price of dispersed storage gets closer to the raw storage price. Comparing to the price of cloud providers, the price of dispersed storage is similar to Amazon S3's price. As the storage size gets bigger and reaches to 1 Petabytes, the price of the dispersed storage system gets very cheaper than the Amazon S3 which is the cheapest provider in that order of size. When the storage size is 10 PB, the dispersed storage costs around 10X cheaper than the Amazon S3 cost. The main reason for cost drop of the dispersed storage is the more cost effective labor cost in the bigger data sizes.

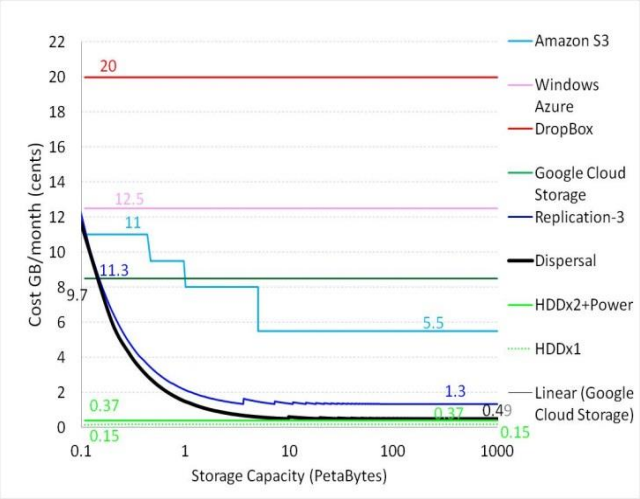

Fig. 3. Cost comparison of storage cloud products with private storage cloud estimations.

We show the breakdown of all costs of our sample private compute cloud comparable to EC2 cc2.8xlarge instance. The pie chart on figure 4 shows that the top 3 most costly parts in the computing cloud are the rental, compute nodes and GPUs. Note that if we consider GPUs are portions of compute nodes then electricity is in the top 3. Therefore power consumption should be seriously considered when building a compute cloud.

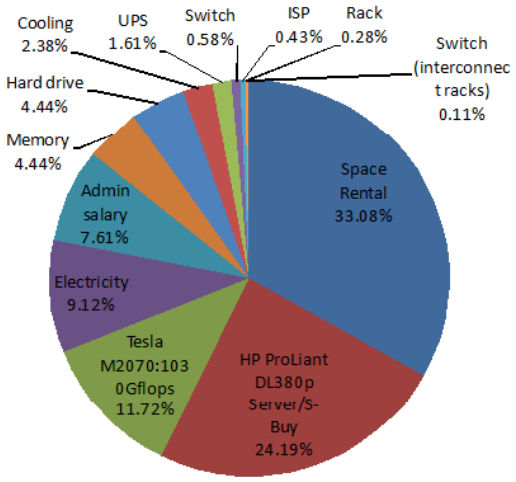

Fig. 4. Cost estimation details of a compute node in a private cloud.

CONCLUSION AND FUTURE WORK

This work estimates the cost of the cloud resources. The comparison between the private and commercial cloud pricing in both storage and compute show promising results. Considering the number of potential users of such systems shows that there is a good opportunity in this area to provide such services in both compute and storage area. The results of the comparison in this project can reveal the potentials in the cloud area, and can be very useful to the new investors in this area.

The future work for this project is to perform a comprehensive performance evaluation of commercial clouds and compare them with the prototype private clouds. The performance evaluation of the current cloud products would give a better view of the accuracy of the comparison. It also determines if the estimated bandwidth of the prototype private clouds is enough to manage large number of the users.

REFERENCES

[1] Hussam Abu-Libdeh et al. RACS: A case for cloud storage diversity. *Proceedings of the 1st ACM Symposium on Cloud Computing*, 229–240, June 2010.

[2] H. Stevens. Cloud Computing Will Be As Influential As E-business. In *Gartner Newsroom*, Online Ed., June 26 2008. http://www.gartner.com/it/page.jsp?id=707508.

[3] I. Raicu, et al. 2011. Making a case for distributed file systems at Exascale. In LSAP '11, 11-18. ACM.

[4] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Cryptology ePrint Archive, Report 2008/489, 2008

[5] Patel C, Shah A; "Cost Model for Planning, Development and Operation of a Data Center", HP Laboratories. http://www.hpl.hp.com/techreports/2005/HPL-2005-107R1.pdf

[6] T. Nufire. 2011 "Petabytes on a budget" from: http://blog.backblaze.com/2011/07/20/petabytes-on-a-budget-v2-0revealing-more-secrets/

[7] Google Cloud Storage. http://code.google.com/apis/storage/

[8] Amazon Simple Storage Service (Amazon S3). http://aws.amazon.com/s3/