Understanding the Cost of Computing in the Cloud

Hina Garg

Rojin Babayan

Vivek Pabani

Illinois Institute of Technology

Abstract

In recent years, most businesses are adopting cloud services at a rapid pace. In fact, a recent 2012 study found that 81% of businesses were either evaluating cloud services, planned a cloud implementation or had already implemented a cloud solution. Organizations realize the benefits that are offered by the cloud. The "cloud" has allowed organizations to avail scalability, flexibility, and agility along with better distribution of workload without significantly increasing IT budget. Though the decision of choice between public and private cloud services has remained to be a dilemma for IT industry. The decision to create a private cloud or use a public cloud is based on various factors. It's important to define and differentiate between private and public cloud services in order to make a well-informed decision on which service to implement. To solve this confusion, we have compared the public and private cloud in terms of its computation power. We have selected seven instance types of Amazon EC2 for our study and done research on the cost effectiveness of public v/s private cloud.

1. Introduction

A public cloud service is provided "as a service" over the Internet and the customer's infrastructure or applications are hosted by a cloud service provider at the cloud provider's premises. Public cloud service is appealing to many organizations as it reduces complexity and long lead times in testing and deploying new applications. It is generally cheaper, too, as there is little or no capital expenditures needed. But there are different types of instances and services provided by public cloud based on the requirements - high computation power, high memory or high end storage. So, the cost of using public cloud depend on the size of cloud and type of service a business want to avail. It may be possible to build a cost effective private cloud equivalent to a public cloud in terms of services offered.

A private cloud service, which can also be called an internal cloud or enterprise cloud, means that the computing infrastructure is hosted on a private platform in the customer data center. It is dedicated to a particular organization and not shared with other organizations. The private cloud resides at the customer's location and offers customers more control over the infrastructure. It is important to keep in mind that a private cloud also offers on-demand capability where more services can be added quickly as needed. the cost of a private cloud can be too high in the initial stage, but it may be cheaper than the public cloud after some point. Our goal is to find such break even point, if it exists. We do this by estimating the cost of a private cloud broken by it's components. The final cost is calculated as cost/GFlops and compared with its public alternative.

2. Private Cloud Components

A cloud consists of different components which attribute to the overall cost. Some components depend on the specific cloud requirements, while others are common for all type of cloud.

2.1. Component Selection and Cost Calculation

The selection of components depend on the requirement of the cloud functions and the size of cloud. The major components can be divided into Hardware, Network, Administration, System Power and Cooling. There are other components as ISP and Location Rental which also have major contribution towards the final cost.

2.1.1. Hardware Components

The hardware for private cloud are chosen to match with public cloud capacity. Amazon uses processors of Intel Xeon family with specific model *Intel Xeon E5-2670* for major instance types. We have selected similar processors keeping the frequency to be same. The number of cores are selected at least as many as provided by public instance type. All the processors are of 1U type. The memory per instance type is selected same as offered by public cloud. Similarly, storage offered by any instance of private cloud is similar to the respective instance of the public cloud. Amazon EC2 offers dedicated SSD with exception of t2-small instance type. The private cloud also offers same amount of dedicated storage per server. The motherboard is selected based on the selected processor and required memory and storage capacity.

2.1.2. Network Components

The network topology used in our configuration is FatTree. We group a number of servers into a rack cabinet using one switch. Then each rack cabinet is connected to another aggregate switch with higher bandwidth capacity. Cost of networking has the following major parts:

- **Network Interface Card:** For those machines whose motherboards do not have an on-board NIC, we need to get a PCI NIC. In particular we need to get Infiniband NICs for the private clouds configured for high quality network instance types (c3.8xlarge, g2.2xlarge, i2.8xlarge, hs1.8xlarge).
- **Rack Cabinet:** We did a research about the size of a rack cabinet. We had two options: either using a smaller rack cabinet with more number of cheaper switches, or using a bigger rack with less number of more expensive switches. Our calculations showed an average reduction of 35% in cost by using bigger racks. Therefore, in each step of the network configuration we used the biggest possible rack cabinet. Moreover, it is also worth mentioning that we decided to leave one or two of the rack units empty for future possible flexibility.
- **Switches:** We used two types of switches in our FatTree topology. First is the rack-mount switch inside each rack to connect all the servers of a rack. And second is the aggregate switch which is more powerful with more bandwidth. For the datacenter of different instance types, we used switches with different bandwidth capacities.

2.1.3. System and Cooling Power Consumption

The hardware equipment used in the cloud setup uses external power for functioning. The power used by each of the hardware equipment varies as per its internal design and functionality. The common equipments are Physical Processor, Memory, Storage, Rack Switch, Aggregation Switch. The total power consumed for establishing the cloud setup can be calculated as:

power consumed = power of equipment (in kwh) * utility cost (per kwh) * PUE

Here PUE can be defined as Power Usage Efficiency. It is calculated as (Total Facility Power)/(IT Equipment Power). A state of the art facility will typically attain a PUE of 1.7 but in usual work environment it is >2 and <2.5. For the provide cloud setup, we are assuming PUE = 2.2 Additionally, out of each watt delivered, about 59% goes to the hardware equipment, 8% goes to power distribution loss, and 33% goes to cooling (Greenberg, et.al, 2009). The utility cost was calculated based on the Commercial utility prices published by US Energy Information Administration. A compounded annual growth rate (CAGR) from the last three years per unit utility prices (from 2012 to 2014) has been calculated. This CAGR was later used to predict the utility cost for each year from 2015 to 2019 (next 5 year period). In this way per unit utility cost has been calculated which is further being used in the "power and cooling cost" calculation.

2.1.4. System Administrators

The administrative cost has been assumed to be approximately $100000 per annum per administrator for managing 500 servers. The number of administrators varies depending on the size of the cloud.

2.1.5. Other Major Components

1. **ISP :** Data and internet link for datacenters are much more reliable, available and with more number of redundancy than typical business links. It is reported that on average a highly available and reliable link for a datacenter is 10 times more expensive than a typical

one. Internet Service Providers often do not release a fixed price for different service types. They often make a contract with the datacenters. It's more cost efficient for both sides if datacenters relay the network traffic from their users to the ISP and based on what they have forwarded, they pay the commission. Unless we have a meeting with an ISP, we cannot find an accurate price for our assumed datacenters. So, we decided to estimate the ISP cost based on two main factors:

- ISP cost often is about 0.5% of the cost of a datacenter (reports confirm)
- Highly reliable and available links are often 10 times more expensive.

The cost of cables router/modem are also included in the cost of ISP.

2. **Location Rental :** The rental of the location depend on the size of the cloud. We have calculated the rent keeping the minimum size of the cloud as 100 servers. The overall rent increases with size of cloud, but the cost of rent per server decreases as we move from small to large cloud.

2.2. Component Details

As described above, the components are selected specific to the instance type. The detailed information is provided in below tables - one for each instance type. The hardware and network components depend on the public cloud capacity of instance type. The admin cost and location rental vary with size of cloud. Power consumption is given for 1 server for 5 year.

Table 1: *Private Cloud Instance Type : t2-small*

Device	**Details**	**Cost**
CPU	10 Core 2.5 GHz	$1534
Memory	2 GB	$49
Storage	None	$0
Motherboard	ATX Single socket H3	$110
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1898.49W. (Cooling 30% of sys power)	$5062
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $2100000
Rental	5 years cost based on size of cloud	$70000 to $3000000
Total Unit Cost	**Cost GFlops with 1 server**	**$26.12**

Table 2: *Private Cloud Instance Type : m3-large*

Device	Details	Cost
CPU	10 Core 2.5 GHz	$1534
Memory	8 GB	$88
Storage	1x32 SSD	$59
Motherboard	uATX Single socket H3	$137
Case	200W 1U Rackmount	$73
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1890W (Cooling 30% of sys power)	$5092
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $3900000
Rental	5 years cost based on size of cloud	$70000 to $3000000
Total Unit Cost	**Cost/GFlops with 1 server**	**$26.13**

Table 3: *Private Cloud Instance Type : c3-8xlarge*

Device	Details	Cost
CPU	2x10 Core 2.8 GHz	$1665
Memory	64 GB	$679
Storage	2x320 SSD	$460
Motherboard	SSI dual socket LGA	$500
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1918W (Cooling 30% of sys power)	$5145
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $3900000
Rental	5 years cost based on size of cloud	$70000 to $2000000
Total Unit Cost	**Cost/GFlops with 1 server**	**$38.53**

Table 4: *Private Cloud Instance Type : r3-4xlarge*

Device	Details	Cost
CPU	10 Core 2.8 GHz	$1534
Memory	128 GB	$2300
Storage	1x320 SSD	$230
Motherboard	ATX single socket LGA	$351
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1932W (Cooling 30% of sys power)	$5180
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $6900000
Rental	5 years cost based on size of cloud	$70000 to $3000000
Total Unit Cost	**Cost/GFlops with 1 server**	**$38.58**

Table 5: *Private Cloud Instance Type : i2-8xlarge*

Device	Details	Cost
CPU	2x10 Core 2.5 GHz	$1534
Memory	256 GB	$4600
Storage	8x800 SSD	$4232
Motherboard	SSI dual socket LGA	$500
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1918W (Cooling 30% of sys power)	$5145
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $3900000
Rental	5 years cost based on size of cloud	$70000 to $2000000
Total Unit Cost	**Cost/GFlops with 1 server**	**$38.87**

Table 6: *Private Cloud Instance Type : g2-2xlarge*

Device	Details	Cost
CPU	8 Core 2.6 GHz	$2296
GPU	Nvidia GTX 680	$499
Memory	16 GB	$168
Storage	1x60 SSD	$57
Motherboard	ATX Single socket H3	$159
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	3712W (Cooling 30% of sys power)	$9950
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $3900000
Rental	5 years cost based on size of cloud	$70000
Total Unit Cost	**Cost/GFlops with 1 server**	**$38.72**

Table 7: *Private Cloud Instance Type : hs1-8xlarge*

Device	Details	Cost
CPU	2x14 Core 2 GHz	$1846
Memory	128 GB	$2300
Storage	24x2000 SSD	$41927
Motherboard	SSI dual socket LGA	$500
Case	200W 1U Rackmount	$78
Rack Unit	48U Rack Enclosure Server Cabinet	$1350
Rack Switch	10 Gbit 48 Port	$5399
Agregation Switch	10 Gbit 24 Port	$3660
Power and Cooling	1957W (Cooling 30% of sys power)	$5251
Administration	100K pa/admin/500 servers	$500000
ISP	5 years cost based on size of cloud	$30000 to $3900000
Rental	5 years cost based on size of cloud	$70000 to $2000000
Total Unit Cost	**Cost/GFlops with 1 server**	**$40.51**

2.3. Cost Analysis

As per our finding, the major components which affect the final cost are administrative cost and power consumption. Both these costs increase linearly with the size of the cloud i.e. number of servers. So, to keep our overall cost low, we tried to keep the number of independent servers low. To achieve this goal, we selected processors with many cores. The more number of cores result in less number of server machines which lowers the over all cost. Other major costs are the ISP cost and the location rental which also decrease when the size of final cloud is small. It should be noted that this cost reduction applies for a cloud of at least 500 servers. A cloud with smaller size than 500 servers will also have to bear the same minimum cost.

3. Cost Comparison of Public v/s Private Cloud

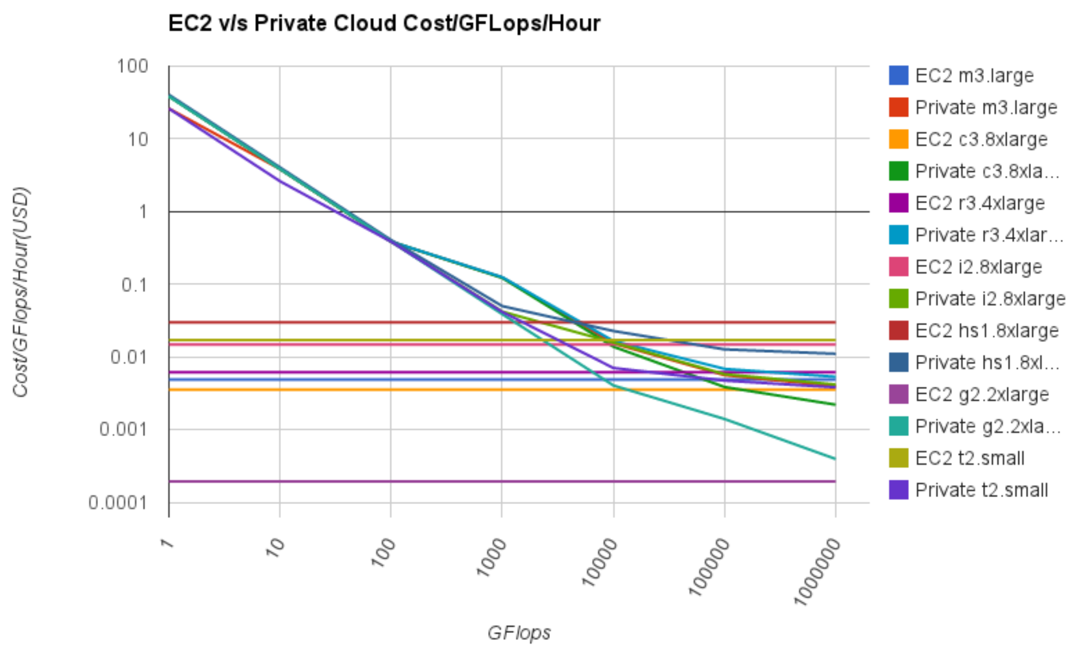

Figure 1: *The cost (in $) per instance type per GFlops per hour, for the 7 Amazon EC2 instance types vs. the 7 private cloud equivalent instance types, from 1GFlop to 1PFlop*

Our study shows promising results for a public cloud in term of cost efficiency. The graph clearly indicates that currently Amazon instances are more cost effective for smaller organizations and businesses. However, on a larger scale, when the number of needed servers increase, setting up a private cloud-based datacenter is a better choice.

We can specify a threshold between 1000 GFlops to 10000 GFlops. If an organization demands a computing power more than this threshold, It is better for it to go for a private cloud option. Nevertheless, history of Amazon instance prices shows that if the trend of reducing Amazon prices continues, we can expect this threshold to go up in future.

Results for t2.small, i2.8xlarge and hs1.8xlarge are more promising for the private cloud because the threshold where cost of private cloud becomes more than the public cloud, is more than other instance types. This indicates that size of the datacenter in term of number of servers plays a more important role in cost estimation of the private cloud than the computing power of the datacenter in term of GFlops. The private t2.small instance type benefit from the fact that it's equivalent public instance provides very low GFlops/instance. The other two have benefit in terms of reduced number of needed servers. Less number of servers means less number of needed administrators, less power consumption and small location, which all result in less expenses. On the other hand, GFlops oriented instance type like g2.2xlarge is always costly in private cloud than the public cloud. The public cloud provides maximum number of GFlops for a relatively small price, which makes it difficult for a private cloud to reach to that level cost effectively.

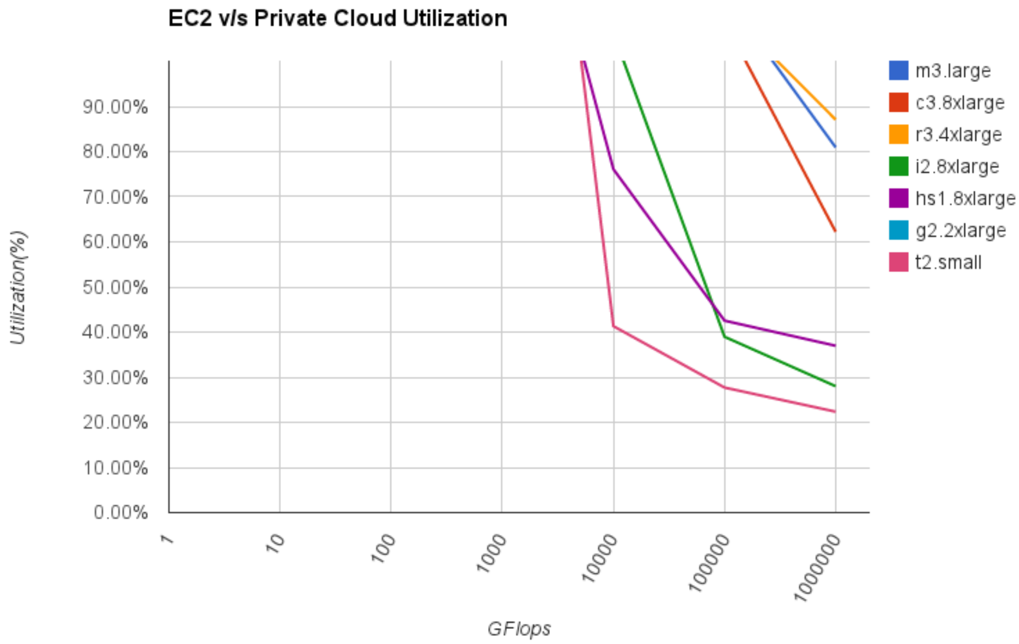

Figure 2: *The utilization of the private cloud from 1GFlop to 1PFlop for the different instance types in order to break even cost wise*

The utilization of the private cloud with respect to the amazon public cloud from 1GFlop to 1PFlop for the different instance types is plotted in the graph. Utilization is calculated as (private cloud cost/ public cloud cost) starting from the break even point. The break even point indicates 100% utilization. As the required number of cloud subscribers (or cloud capacity in GFlop) increases, there is more cost saving using a private cloud v/s a public cloud after the breakeven point. For example while using private cloud setup in t2.small instance, we are incurring only 41% of the public cloud cost when the GFlops reaches around 5000 (i.e. between 1000 and 10000). This cost reaches to 22% when the GFlops reaches to 1 million. In other words, A cloud with high number of users is cost effective even when it is used for only 30-50% of its full capacity.

The utilization value changes with the instance type. Since we are comparing the public and private cloud in terms of the cost/GFlops, the utilization value depend on the GFlops provided by the public instance instead of the size of the public instance. For example public t2.small instance is the cheapest in terms of cost/hour, but it is the costliest in terms of cost/GFlops among the seven instance type chosen for this project. Due to this fact, its private alternative provides the best breakeven at approx. 5000 GFlop. On the other hand, public g2.2xlarge instance, being one of the costliest among the seven instance types, provides the maximum GFlops for a very low rate. We are considering the single precision flops for GPU which is about 3TFlops/sec. Due to this, we cannot achieve breakeven for private g2.2xlarge instance. All other instances have utilization values in between these two instances. The utilization values are almost in the same order as the public cost/GFlops value of seven instances.

4. Conclusion

The research and analysis of this project solves the confusion of selecting a private cloud v/s public cloud. As suspected in the beginning, it is possible to build a cost effective private cloud which offers similar services as the public cloud. This applies when the requirement in terms of total computing power is high. Using this study, a mid size or large size organization can decide whether to build a private cloud or rent the public cloud like amazon based on their requirement and number of users.

5. Contribution

1. **Hina Garg** [A20342375] [hgarg@hawk.iit.edu]
    - Research on system and cooling power consumption and administrative cost.
    - Design and explanation of Plot#2.
2. **Rojin Babayan** [A20334414] [rbabayan@hawk.iit.edu]
    - Research on network and ISP components.
    - Design and explanation of Plot#1.
3. **Vivek Pabani** [A20332117] [vpabani@hawk.iit.edu]
    - Research on hardware components.
    - Design of Plot#1 and Plot#2.
    - Design of tables for 7 private instance types.