The Method and Tool of Cost Analysis for Cloud Computing

Xinhui Li, Ying Li, Tiancheng Liu, Jie Qiu, Fengchun Wang
IBM China Research Lab, BJ, 100193, China
{lixinhui, lying, liutc, qiujie, wangfc}@cn.ibm.com

Abstract

Proposal of Cloud Computing is tightly coupled with low cost. Reduction of cost is considered as an important advantage of Cloud. However, there are no available tools proper for cost calculation and analysis in Cloud environment. This paper presents our efforts towards filling in the gap. We format suits of metrics and formulas for the calculation of Cloud Total Cost of Ownership (TCO) and Utilization Cost, considering the elastic feature of Cloud infrastructure and widely adopted virtualization technology in Cloud. This provides a foundation for evaluating economic efficiency of Cloud and provides indications for cost optimization of Cloud. We have developed our calculation and analysis approach into a web tool which is used in the internal Cloud environment and demonstrate initially its analysis capability on the cost distribution and utilization imbalance factor.

1 Introduction

Cloud computing thrives today with the potential to transform a large part of the IT industry. Cloud computing refers to both the applications delivered as services over the Internet and the servers and system software in the data centers that provide those services [2]. The datacenter servers and software are what we call Cloud in this paper. Cloud adopts an on-demand infrastructure to provide its computing resource as an elastic service. Existing technologies highlight the benefits cloud users can gain. Cloud users can apply for resources from Cloud and then return back to cope with their business needs. By the pricing model of paying-as-you-go, Cloud users only pay for the resources allocated to them. This can benefit Cloud users by transferring the risk of economic losses caused by over proportioning (under utilization) and under-provisioning (saturation) to Cloud providers. However, few sees Cloud from the perspective of cloud provider. How much economic risk will Cloud providers take?

The Cloud is expected by its provider to offer services below the costs of private datacenter and yet still make a good profit by taking full advantage of economies of scale, combined with statistical multiplexing to increase resource utilization. When more and more data center owners, triggered by the expected good return, rush to operate their own Clouds, the lack of approaches and tools to analyze the economic efficiency of these practices results in a weak foundation for the risk evolution and cost optimization of Cloud. Can the cost be a cardiac stimulant to the Cloud at all? The economic efficiency becomes crucial for cloud providers.

The attainment of economic efficiency in a Cloud can be treated as a two-step process. Firstly, the resource must be applied in an area in which its benefits outweigh its costs. Secondly, for maximum efficiency to be achieved with given resource, it must be applied in such a way that the costs are minimal. In this way, cost analysis is the exact foundation to economic efficiency analysis. For this purpose, two major costs of Cloud are required to be understood and analyzed: Total Cost of Ownership (TCO) and Utilization Cost.

Total Cost of Ownership (TCO) is generally used as a means of addressing the real costs attributing to owning and managing an IT infrastructure in a business [14]. TCO not only includes the capital cost, but also the cost of operating the IT infrastructure. The comprehensive consideration on the whole lifetime spending makes TCO proper to providing a cost basis for determining the economic value of Cloud investment. So we format the calculation of Cloud TCO as the basis for Cloud cost analysis.

Higher work output against less expensive resource consumptions is the key to the development of Cloud, which is especially meaningful to achieve maximum economies of scale. For this purpose, the utilization of different resource coping with users' dynamic demand is critically important. For Cloud, it is not enough to only calculate and analyze the TCO. The cost directly associating with the real resources locked up or committed to a particular user or application should also be analyzed, which is referred to Utilization Cost.

Although there have been some technologies which are related to the cost analysis of traditional data centers, the characteristics of Cloud make them difficult to be adopted for the Cloud cost analysis.

Elastic resource utilization: Cloud provides an on-demand resource service that users can apply for resources from Cloud and then return back to cope with their business needs. Underlying, Cloud adopts the architecture which can continuously adapt to the users' changing requirements automatically. Dynamic scalability technologies support Cloud to operate a resource pool in N, N+1, or 2N configurations without human intervention, which means that servers, software, power and facilities including UPS and battery system, can be added into or removed from the resource pool in a manner of coping with the change of users' demands. This makes the cost analysis of Cloud completely different from traditional

978-0-7695-3840-2/09 $26.00 © 2009 IEEE
DOI 10.1109/CLOUD.2009.84

datacenters'. Cloud cost analysis should be utilization-oriented, changing constantly to cope with the changing requirements of users, furthermore, the user scales.

Current cost analysis technologies do not consider the elastic characteristic of Cloud that their calculation mainly depends on statistics of the each cost item and summary of all the cash outlays, which neglects the impact of elastic utilization on cost calculation. Either, current technologies can not execute calculation without the exact data of every cost item which is difficult to get for the always changing Cloud.

Wide adopted virtualization: In the Cloud, multiple applications may be hosted on a common set of servers. This allows for consolidation of application workloads on a smaller number of servers that may be kept better utilized, as different workloads may have different resource utilization footprints and may further differ in their temporal variations. For this purpose, virtualization technologies are widely adopted in the Cloud. Cloud providers package arbitrary user applications as sets of virtual machines (VMs) and provide such VMs as the resources for applications.

In this way, VM becomes the unit of resource in Cloud, which will definitely impact the cost analysis. For example, traditional cost calculation mainly takes physical servers as the unit to calculate servers cost and refreshes should be included. But after the adoption of virtualization, the usage of physical servers is consolidated. Another example is software. Software in Cloud is licensed and used in the image of VM for cloud users, which requires calculating the needed software license by VM rather than physical server. Cloud providers loan software to users during a lease of VM applying, which make software in Cloud share in the long period and software cost impractical to calculate directly on the exclusive price as does the traditional accounting.

In this paper, we format and calculate both TCO and Utilization Cost of Cloud, aiming to provide the foundation for evaluating the economic efficiency of Cloud: Firstly, a suite of metrics and formula are introduced to express and quantify Cloud TCO. Secondly, connections among the utilization of Cloud and the cost metrics are analyzed and used to calculate the utilization cost of Cloud. When Cloud users change their resource requirements, costs on software, physical servers, facilities, and other cost items change accordingly. Thirdly, a tool to calculate and analyze Cloud TCO and Utilization Cost has been developed as a web service for convenient usage.

Based on the internal cloud environment in IBM China Research Laboratory, we used the approach and tool to do the initial cost analysis. We will be engaged to put the cost analysis tool in the practice of the customer' s Cloud environments and try to draw guideline rules from our analysis for realizing Cloud practices with high economic efficiency.

The rest of this paper proceeds as follows. Section 2 introduces a suite of metrics used to express and quantify Cloud TCO. Section 3 describes our approach to calculating Cloud utilization cost. Section 4 demonstrates the usage of our cost analysis tool and the effect of calculation. Finally, Sections 5 and 6 discuss related work and conclusions.

2 Calculation of Cloud TCO

Cloud TCO is referred to as the cost spent to build and operate a Cloud. The items composing the cost of Cloud have been collected, and we separate these items into eight categories as: server, software, facilities, support and maintenance, network, power, cooling, and real-estate. Although the categories look common, the Cloud characteristics are considered inside for measurement and calculation of each specific cost category.

2.1 Cloud Cost Amortization Model

Because of different purchase prices and duration over which purchased resources are planned to use, cost of different metrics items can not be put together directly or comparative fairly. Furthermore, resources of Cloud are provided to users in a paying-as-you-go model. Then it is necessary to analyze how the costs of server, power, and cooling and other metrics item contribute to the monthly renting. For this purpose, the monthly depreciation cost (or amortization cost) is calculated for each metrics item from the initial purchase expense based on the duration over which the investment is amortized (which is related to its expected lifetime) and the assumed interest rate. Especially for the items contributing operational expense like power, cooling, and maintenance, amortization is as easy to calcite their monthly duration pays. Typically, real-estate is depreciated over periods of 10 years and other items like server and facilities are depreciated over three years with 5% money cost [29]. By amortization, we can get a cost amortizable rate parameter, Arp, which we can apply to both one time purchases (server cost, facility cost and etc.) and operational expenses (power cost and cooling cost etc.).

Table1. Parameters for Amortization:

Para.	Description
A_p	Amortization period unit (year / month / hour, default as one month).
time	hours consumed

$Arp = (1+0.05) * time / (30*24*A_p)$; (1)

2.2 Calculation Model

Server Cost: In Cloud, servers are mounted in the racks and constructed into a resource pool. Resources are allocated from this pool to users. Calculation of TCO takes all these servers into consideration. We assume that all the servers with same type and the same CPU, memory, disk and any other servers' configuration because almost all of the current clouds are made up from homogeneous machines.

Table2. Parameters for servers cost

Para.	Description
N_s	Number of physical servers in resource pool
VI_{ps}	Cost per physical server of same configuration

$Cost_{servers} = VI_{ps} * N_s * Arp(time)$; (2)

Software Cost: This cost is mainly caused by the licenses should pay for, According to how software is licensed, the software is divided into three types. The costs these types are calculated differently: Type I software include Operating Systems and any other software being licensed by suite and the license number of the software of this type equals to the number of virtual images applied by Cloud users. Type II software are Application Server, VM software and any other software that are licensed by the processor number and the license number is always calculated as the number of processors where these software run. And Type III software are mostly management software which are licensed by the processor number of the servers that are managed by the software, and the license number is calculated by the processors they manage. We use the following table to describe the parameters required and formula for software calculation.

Table3. Parameters for software cost

Parameter	Description
VI_s	Unit Price of Type II software
VI_o	Unit Price of Type I software
VI_m	Unit Price of Type III software
S_s, S_o, S_m	Subscription factor – percentage of unit price that yield annual fee
$N_{serverlic}$	Number of Type II software license
N_{oslic}	Number of Type I software license
$N_{monitorlic}$	Number of Type III software license

$Cost_{software} = [S_s *VI_s* N_{serverlic} + S_o *VI_o* N_{oslic} + S_m *VI_m * N_{monitorlic}] * Arp(time)$; (3)

Network Cost: Cost related to networking is caused by switches, NIC and cables which are used to attach physical servers to the network. For the TCO calculation, the NIC requirements are accounted for in the price of the physical servers, since they often accompany server purchase. Therefore, this section concentrates only on the network switches. Following table describe the parameters and formula for network cost calculation.

Table4. Parameters for networking cost

Para.	Description
N_{switch}	Number of new network switches per year
S_{NIC}	Number of NIC per virtualized server
P_{NIC}	Number of ports per NIC
P_s	Price per switch
N_{port}	Port number of a network switch

$Cost_{networking} = P_s * N_{switch} * Arp(time)$; (4)

$N_{switch} = S_{NIC} * P_{NIC} * N_s / N_{port}$ (5)

Support and Maintenance Cost: This metrics item also belongs to soft cost, but include some important work, such as software distribution and upgrading, asset management, troubleshooting, traffic management, servers' configuration, virus protection, disk management, and performance maintenance. Here we can give an overall formula to calculate this part of cost which is calculated by labor efforts. The following table describes the parameters needed.

Table5. Parameters for labor cost

Para.	Description
N_{Labor}	number of administrators responsible for support and maintenance
T_{use}	Average time spent on unit system under utilization
T_{idle}	Time spent on all the idle systems
R_{Salary}	Rating number of salary averages

$Cost_{support\ and\ maintenance} = N_{Labor} (T_{use} * N_s + T_{idle}) R_{Salary}$; (6)

Power Cost: Power consumption in Cloud is primarily caused by Computing Infrastructure (IT loads) including Server servers, network switches, and Network Critical Physical Infrastructure (NCPI, non-IT loads): transformers, uninterruptible power supplies (UPS), power wiring, fans, air conditioners, pumps, humidifiers, and lighting [4]. Because they are plugged into a rack as components and configured with the increment of racks, the power consumed by these items can be calculated by the unit of rack.

Table6. Parameters for power cost

Para.	Description
S_{rp}	Sum of the power rating of working servers (it can be gotten from website of servers vendor)
E_s	Price per hour of 1kw of electricity
L_s	Steady-state constant
N_{rack}	Number of racks in working

$Cost_{power} = L_s * E_s * S_{rp} * N_{rack} *Arp(time)$; (7)

Cooling Cost: As a first-order approximation, the power consumed in the data center is completely converted to heat. Therefore, the power rating in Watts of rack is equal to thermal output [5]. We define a parameter L representing how much that power is consumed by the cooling equipment for every 1W of heat dissipation in the data center. L is statistically 0.6 according to experiments completed in our laboratories. This number is confirmed by Forrester Research, Inc., which estimates that 0.5W to 1.0W of power is required to dissipate 1W of heat [10]. The details of parameters and formula are under below:

Table7. Parameters for cooling cost

Para.	Description
L	Cooling Load Factor - amount of power consumed by the cooling equipment for 1W of

	heat dissipated [3]
P	Airflow Redundancy Constant – redundant airflow required to cool data center [1]
H	Inefficiency (Humidification) Constant – redundant airflow to account for burden of humidification [1]

$Cost_{cooling} = L * (1+P) * \mathbf{PowerCost}(time) / H$; (8)

Facilities cost: Facilities are not equipment but necessary for the normal running of equipments, like PDU, KVM, cables and etc. They are all wrapped into racks. The parameters and formulas are as follows:

Table8. Parameters for facilities cost

Para.	Description
N_{rack}	Number of racks
VPfp	Price of facilities per rack
time	Hours consumed

$Cost_{facilities} = N_{rack} * VPfp * Arp(time)$; (9)

Real-Estate Cost: Due to the special infrastructure required like cooling and power, space needed by a Cloud is often significantly more expensive to build than standard commercial properties. A data center rated at 40W per square foot costs approximately 400 dollar per square foot [8], we get the total space of data center according to rack number. The parameters and formulas are listed below:

Table 9. Parameters for space cost

Para.	Description
A_p	Cost per square foot to build Cloud [8]
R_{SF}	Square feet per rack
R_{SPACE}	Percent of space taken by racks in all(<1)
W_{Server}	Weight of a physical server
W_{rack}	Weight of a rack
$C_{presssure}$	Constant pressure confronted by unit floor
R_a	Annual Percentage Rate

$Cost_{space} = A_p * S_{pace} * Arp$; (10)

Where, S_{pace} is space taken up by all the racks under utilization. It can be gotten from following formula.

$S_{pace} = (R_{SF} * N_{rack}) / R_{SPACE}$; (11)

Moreover, the value of pressure confronted by unit floor can not beyond the $C_{presssure}$.

$(N_s * W_{Server} + N_{rack} * W_{rack}) / S_{pace} <= C_{presssure}$; (12)

Cooperation of formula (11) and (12) can get a minimum S_{pace} which can be used in formula (10) for space cost.

With this annuity factor, cost calculated from formula (10) is the real-state cost paid per year.

TCO: The sum of all the eight kinds of cost is the Cloud TCO. When used for calculation, the parameters required by these formulas can either be input by Cloud operators or collected from industry statistics.

3 Calculation of Cloud Utilization Cost

Although TCO is absolutely helpful to evaluate IT cost of a Cloud during its whole lifetime, it is more apt to be used to evaluate the foundational cost instead to reflect elastic resource delivery of cloud. By the foundational cost, we mean the part of cost which does not change a lot with the utilization of Cloud by users. TCO consists of the cost of all the servers in resource pool, all the facilities to support these servers and so on. However, Cloud just uses part of these servers and resources of other categories to satisfy the users' workload. Cost of the used part of resources, changing sensitively with various work loads, is important to know the usage efficiency of given resources. We refer to the cost caused by users' utilization as the Utilization Cost and want to format its calculation. Our format takes the characters, virtualization and elasticity, of Cloud into consideration:

Firstly, virtualization is widely adopted in the Cloud and Cloud providers package arbitrary user applications as sets of virtual machines (VMs) and provide such VMs the resources that these applications need. In other words, the basic unit of Cloud consumed on behalf of users' requests is not the physical server but VMs users applied instead. So we adopt VM as the inputs of our calculation model.

Secondly, the elasticity of cloud prevents the calculation methods depending on monitoring the usage of all kinds of resources improper here. An atomic effort is needed on monitoring different sides of changes with changing users' demands, if it is not impossible, from the utilization of pooled servers, all kinds of operational expense, to each cash outlay spent on increasing new equipments and facilities into the resource pool when existing pooled resources can not satisfy users. So we adopt a derivation-based approach which calculates the Utilization Cost from input VMs without any dependency on monitoring or accounting.

3.1 Calculation Model

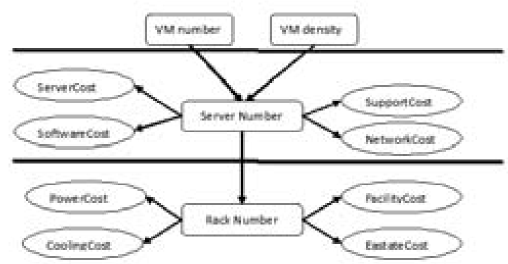

Fig.1. Three-layer model to calculate utilization cost

Completely different from traditional cost accounting methods, we adopt VMs as the inputs and use a three-layer derivation model to calculate the utilization cost considering amortization, illustrated in the figure 1. Utilization cost is a subset of TCO. The same eight metrics items, i.e. Server Cost, Software Cost, Support Cost, Network Cost, Power Cost, Cooling Cost, Facility Cost and Real-estate Cost, compose the utilization cost. The overall calculating process is: First of all, we get VM number and VM density as input in layer one. Then in layer two we calculate Server Number according to these two parameters. Server Cost, Software Cost, Support Cost, Network Cost can be got on the basis of Server Number and VM number. In layer three, we compute the number of racks which contain servers. And the Power Cost, Cooling Cost, Facility Cost and Real-estate Cost are rest with rack number. At last, we sum these eight costs up to get utilization cost of Cloud. The details of calculating are under below:

VM Density and Number of Physical Servers: VM density is the number of VMs can be hosted on one physical server. It depends on the capacity of hosting servers, the resources required by each VM, and the Hypervisors' management capability of how much the total virtual CPUs, virtual memories or any other virtual resources can be allocated to VMs. The report of TANEJA group [30] presents the VM density values of two primary Hypervisor softwares, VMware and Microsoft Hyper-V. Given N_{vm} representing the number of applied VMs, following formula is used to get N_s, the number of physical servers:

N_s = ceil [N_{vm} / VM density]; (14)

Cloud providers can use VM density provided by some VM vendors and servers producers. But for better customization and accuracy, we suggest Cloud providers making their own test for the value of VM density based on the real Cloud and the applications.

Number of Racks: As we investigate, 42U standard rack is the most widely used rack, and 40 percent of the space is used for KVM [8], facilities and etc. As a result, if we get the space a server needs and the server number Ns, we can calculate the number of racks acquired for providing services. For example, one IBM X3950 server needs 4U to be contained, so the servers' number of a rack is six (42U*0.6/4U). The parameters and formula are under below:

Table10. Parameters for Rack number

Para.	Description
V_{rack}	Units taken by per server
UV_{server}	Units available of a rack for servers

Based on the derivate number of physical servers, we can get minimum number of racks to contain these servers by adopting Bin-Packing Algorithm [28].

N_{rack} = Function of Bin-Packing (N_s, V_{rack}, UV_{server}); (16)

Utilization Cost: By inputting the derivate number of racks and physical servers into Formulas introduced in section 2, eight aspects of cost are gotten and their sum will be the utilization cost, which dynamically copes with the users' demands. Followings are more details:

A given number of equipments require a matched number of facilities to support. One full capacity X3950 needs 3 cables, 1440 watts/UPS and 5527 BTUS. In this way, the number of facilities can be derived from their matching relationship with given physical servers. For each different type of physical servers, their mapping relationship with facilities is different and kept separately for the reference of calculation.

3.2 Discussion

This mechanism provides a convenient way to derive the utilization cost. Cloud providers can also input derivation rules specific to their own infrastructure or detailed deployment in their Cloud. We provide two ways for derivation rule inputs: one is to assign the relationship between two different metrics items. The other is to assign the relationship between the costs of two different metrics items. This helps customize our three-layer model easily with different Clouds.

From the mention referred above we can indicate that all the aspects of utilization cost such as server cost, software cost, power cost, real Estate Cost and etc is mainly determined by the physical servers number, which represent the minimum number of servers to host VMs without any disruptive performance impacts. For example, the VM density is assumed to be six, and the utilization cost of hosting VMs ranging from one to six in one server are all the same. As a result, if the server cost increase directly, other cost will increase simultaneously in all. Furthermore, this also provides constraints for TCO calculation model because the server number derived by this mechanism will be the minimum number of servers required to satisfy users' demands.

4 Calculation Tool

We have developed a cost calculation and analysis tool based on our format to Cloud TCO and utilization cost. This tool provides both web page user interface and web service interface. Users can invoke this web service for the function of Cloud TCO and Utilization Cost calculation. Besides the web service, we also provide a web tool considering better user experience. At present, limited analysis has been done on the imbalance factor of different cost metrics and cost distribution. In future, our focus will be tuned to this. In the following sections, we will demonstrate through the cost calculation web page interface and present some interesting analysis result. .

4.1 Calculation of TCO

We have introduced the metrics and the parameters needed to calculate them in previous sections. Those parameters are relatively statistic. For example, the per-processor price of an Application Server is applicable to multiple calculations for a period of time. This price can be input by user and used for the calculation of software cost till next update. Fig.2 presents screenshot of the web-page for metrics parameter tree. Each branch represents a cost metrics item. Each child label on the branch represents a parameter under the given cost metrics item.

Users can click on any child label of the tree and fill into popup text field for assigning value to this parameter. After users input the values, the validation work is made according to the constraints from the derivation rules of section 3. If the input value can not be satisfying, derived value will be used in real calculation. For example, input rack number is less than the number derived which is the minimum number needed to match servers. Then the derived rack number will be used in calculation. Derivation can free users from the pressure to gain all the accurate inputs, which is especially useful to avoid error or unreasonable value assignments in the case that users are not sure on some parameters. For some parameters relatively steady for characters of industry, we also provide defaults by industrial survey. After the parameter assignment, calculation task can be submitted to get TCO.

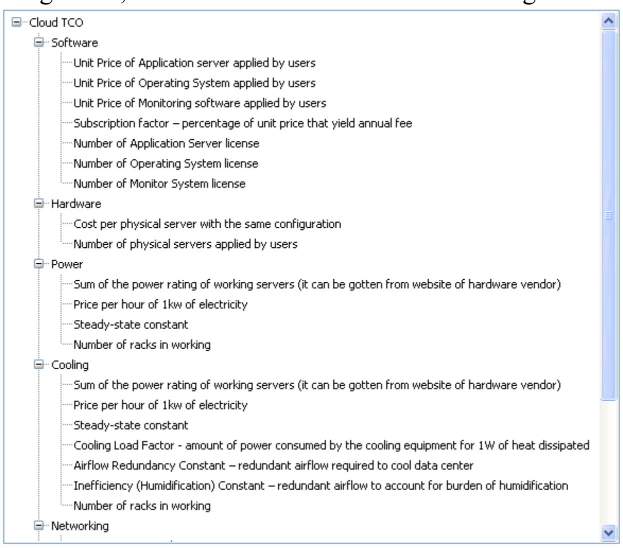

Fig.2. Metrics tree for TCO calculation

4.2 Calculation of Utilization Cost

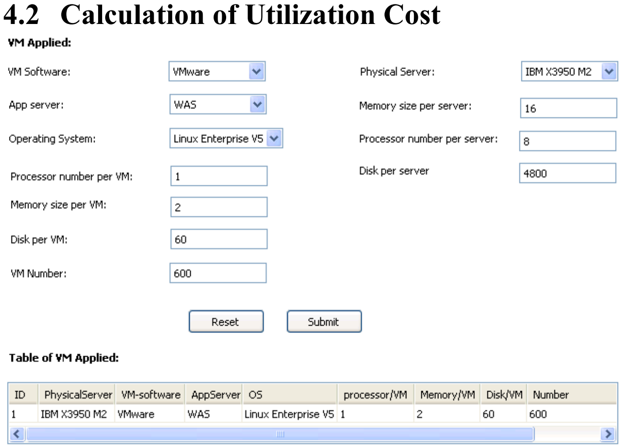

Fig.3. Input the VM information applied by Cloud users

Utilization cost is tightly related to the utilization status of Cloud at the time of calculation. Fig.3 presents a screenshot for Cloud providers to input the running VMs at the time and concrete configurations of these VMs, such as the selection of application server, and processor number per VM.

For simplicity, 600 VMs of the same configuration is input on this page as a demonstration. After inputting, a table on the bottom of this page will be presented with a new item of VM information of the input type.

According to the formula of Section 3, the mutilation cost can be calculated from the inputs of running VMs.

4.3 Output after Calculation

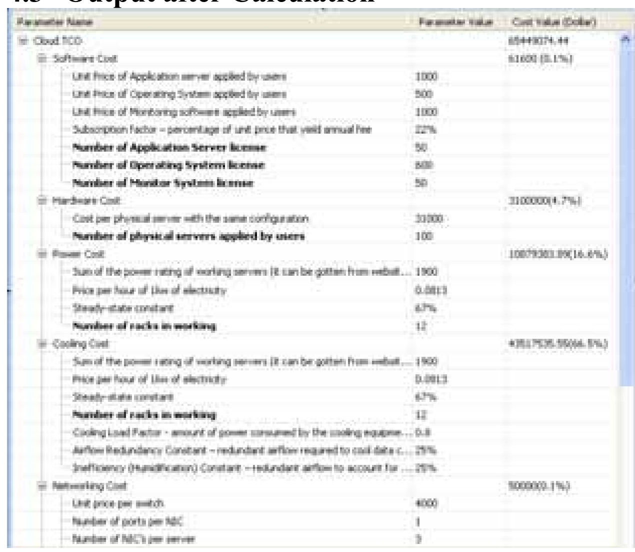

Fig.4. Output of TCO calculation

Fig.4 is the output screenshot. The presented tree-table includes the parameters input and cost value resulted from our TCO calculation. The first column of the table is still the metrics tree. Some labels in **bold font** represent that value of these parameters are not input by users but gotten from the derivation described in section 3. The second column is values of the parameters, input or derivate. In this case 600 running VMs, 100 physical X3950 servers and 12 racks are required. Number of software license and faculties are accordingly calculated. The last column presents the cost value of each cost metrics and the percentage taken by the metrics item on the total cost presented in the field of Root node.

Users can moderate the parameter values for comparison among several configurations. Each time a single parameter is modified, a chain of reactions on other parameters' value is automatically triggered by derivation rules and the cost column will reflect the changes. It provides flexible way to analyze the impact of different parameters on the final TCO. From the data presented on Fig.4, servers, po wer, cooling, and labor costs are the most primary contributors to Cloud TCO.

Output of utilization cost looks similar since just the inputs are utilization status and the utilization cost is gotten from derivation.

4.4 Analysis on Balance Factor

We have made some initial cost analysis and here present an analysis of balance factor as an example of our tool's usage. A high rate of utilization at minimum cost is attainable for the Cloud as a whole only if it is balanced (if all cost items are utilized to about the same extent). If this is not the case, those more heavily utilized items will potentially form bottlenecks. These bottlenecks will preclude the attainment of higher rates of utilization for the under-utilized items. Balance factor is used to describe the balance extent of all cost items' utilization and is very important to indicate where potential bottlenecks are being developed. An increase in the capacity of bottle-necked item could defer saturation for some time. We also find the recommendation to add bottlenecked items inducing the utilization increase of other items. Doubling the capacity of bottlenecked facilities, for example, will not definitely lead to a doubling of the whole Cloud capacity. It can also change the cooling utilization accordingly. So it is required not only to calculate the balance factor but also to identify and consider the connected utilization changes among various items, which can be easily achieved by our two cost calculation models.

We evaluate the utilization parameter of each cost item as the percentage value that the number of Utilization Cost on this item takes in the number of Cloud TCO on the same item. For example, the server utilization is the percentage of server's utilization cost takes in the overall cost of pooled servers. Furthermore, we define standard deviation of the utilization parameters of different items as the Cloud balance factor. The factor varies between 0, for the perfectly balanced cloud, and 1, for maximum unbalances cloud. The bigger it is, the more likelihood that bottleneck will emerge in a given cloud.

In our internal cloud environment, we own 16 X3950 servers, 10 RedHat linux OS, 8 WAS and 8 VMware Exsi per server according to the PVU[14] calculation methodology and estimated maximum number of VMs can host on a server. Followings are the analysis to our experimental result.

This histogram of Fig.5 presents the utilization parameters of various cost items. The histogram bars represent the utilization parameter of seven various cost items and the area above the bars represents slack in the Cloud. From their heights, the balance factor is derived as 0.2126, which means that all items are not utilized to about the approximate extent. The figure indicates the potential bottlenecks are with facilitates. After doubling the facility capacity, we get the Fig.6. It does not present any improvement on the balance factor. Instead, the balance factor becomes worse (0.2818). And the doubling of facilitates can not double the Cloud capacity. The whole capacity increases from 70 VMs up to 120 VMs which is only 70% improvement instead of 100%. Clearly server is the next bottleneck. Because our calculation of Utilization Cost already considers the tight relationships among different cost metrics, we can present chained changes caused by the change of one single item in the utilization parameters of other items and balance factor accordingly. By using our tool, the users can do several trials to change different parameters to finally figure out a best fit solution for increasing the Cloud capacity with a economy effective way.

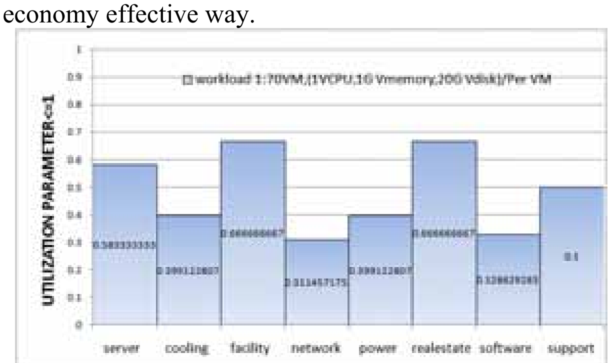

Fig.5. Histograms with a balance factor of 0.2126

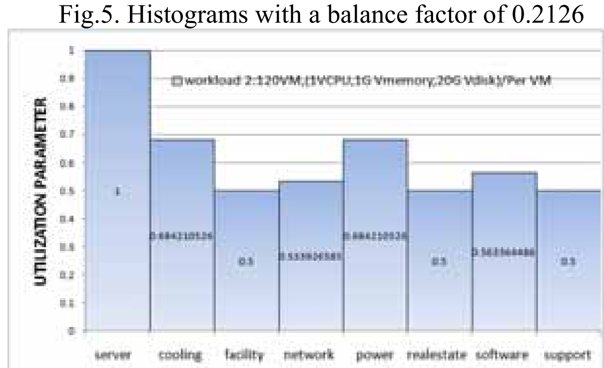

Fig.6. Histograms with a balance factor of 0.2818

5 Related Work

TCO was firstly developed in 1987 [14]. Besides cost calculation, TCO can also be used to evaluate and weight multiple solutions available for one specific problem. In a framework of Brocke, TCO and ROI concepts are introduced in the corporate level for evaluating the efficiency of investments in the design of the information system architecture [15]. Danilo presents an approach to treat complex infrastructural design issues as a single cost-minimization problem [16]. However, TCO calculation in these researches is given in conceptual level, which is in means of calculating the monetary profit of the investment gradually.

Currently, there are also some industrial tools for TCO. Company Lounge provides a tool for the selection of servers [17]. Its calculation has special focuses on server cost. HP provides a tool for a TCO comparison between Blade and Rack-mount server [18]. Alinean's tool [19] is designed to evaluate their Data backup solution with focus on Data transferring cost. PMTC focus on networking deployment cost [20]. Solarrcom focus on the evaluation of desktop devices [21]. These tools have specific focus on some metrics but give other parts rough estimates. APC give series calculators for Data Center on the conventional Legacy architecture lifecycle costs, on the energy cost and carbon footprint, and the power related capital [22, 23, 24, and 25]. These tools do not consider virtualization or elastic utilization. Tool in [26] focus on the consolidation effect of virtualization on energy cost but does not consider virtualization in the whole definition of TCO metrics. These tools are based on accounting and summing every cash outlay. These technologies don't consider the dependencies among cost metrics. None of them consider the format of Utilization Cost calculation.

6 Conclusions and Future Work

Elastic resource utilization and widely using virtualization in the Cloud challenge existing cost analysis methods. We identify two important aspects, Cloud TCO and Cloud Utilization Cost, to evaluate the economy efficiency of Cloud. Cloud TCO embodies the investment. Cloud Utilization Cost embodies the cost caused by the Cloud utility and reflects sensitively the dynamic utility. Both of them are very meaningful to Cloud cost analysis. In this paper, we express and quantify cloud TCO with a suite of metrics and formulas considering the characters of virtualization and elasticity. Utilization Cost leverages the metrics items defined in Cloud TCO and takes workload, i.e. running VM number, as input. Further more, dependencies among different cost metrics are identified and used in the calculation of Utilization Cost. We develop these calculation models into a tool which can execute cost calculation without knowing exact data of every cost item and provides a flexible way to analyze the effect of different metrics on the cost. In the internal Cloud environment in IBM China Research Laboratory, we have used the tool making some initial cost analysis. In this paper, we present an analysis example on the balance factor of Cloud.

We will be engaged to put the cost analysis tool in the practice of the customer' s Cloud environments and try to draw guideline rules from our analysis for realizing Cloud practices with high economic efficiency.

7 References

[1] McFarlane, Robert, "Let's Add an Air Conditioner," SearchDataCenter news article, published November 30, 2005. http://searchdatacenter.techtarget.com/columnItem/0,294698,sid80_gci1148906,00.html

[2] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Above the Clouds: A Berkeley View of Cloud Computing, University of California Electrical Engineering & Computer Science, February 10th, 2009.

[3] Patel, Chandrakant D., Shah, Amip J., "Cost Model for Planning, Development, and Operation of a Data Center," Internet Systems and Storage Laboratory, HP Laboratories, Palo Alto, June 9, 2005

[4] Sawyer, Richard, "Calculating Total Power Requirements for Data Centers", American Power Conversion, 2004

[5] Rasmussen, Neil, "Calculating Total Cooling Requirements for Data Centers", American Power Conversion, 2003

[6] "CustomerCaseStudies - OpExSavings.xls" (INTERNAL USE ONLY)

[7] Disaster Recovery. (2006, July 13). In Wikipedia, The Free Encyclopedia. Retrieved July 17, 2006, from http://en.wikipedia.org/wiki/Disaster_recovery

[8] Anthes, Gary, "Data Centers Get a Makeover", Computerworld news article, published November 1, 2005. http://www.computerworld.com/databasetopics/data/datacenter/story/0,10801,97021,00.html?SKC=home97021

[9] Hughes, Ron, "The data center of the future-Part 1-Current trends", The Data Center Journal news article, published May17, 2005.

[10] Fichera, Richard, "Power And Cooling Heat Up The Data Center", Forrester Research, Inc. March 8, 2006

[11] Dunlap, Kevin, and Rasmussen, Neil, "The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers", American Power Conversion, TCO

[12] Vouk, M.A. Cloud Computing - Issues, research and implementations, IEEE Information Technology Interfaces 30th International Conference, page(s): 31~40, June, 2008.

[13] VMware. Consolidating Web Applications Using VMware Infrastructure. 2008.

[14] Gartner Total Cost of Ownership. http://amt.gartner.com/TCO/index.htm

[15]JV Brocke, MA Lindner, Service portfolio measurement: a framework for evaluating the financial consequences of out-tasking decisions, Proceedings of the 2nd international conference on Service oriented computing, 2004

[16]D Ardagna, C Francalanci, M Trubian, A cost-oriented approach for infrastructural design, Proceedings of the 2004 ACM symposium on Applied computing, 2004

[17] SizingLounge. http://www.sizinglounge.com/index.php?type=sizing&gclid=COfz_bfTgJgCFQEMewodID7BDA

[18] HP Blade and Rack-Mount Server Cost Comparison Calculator. http://www.itbusinessedge.com/offer.aspx?o=00190119rc

[19] Alinean,Inc. BakBone TCO Calculator. https://roianalyst.alinean.com/roi_calculators/AutoLogin.do?d=208178774968639125

[20] PMTC. http://sourceforge.net/project/showfiles.php?group_id=74164

[21] SOLARCOM. TCO calculator of IT Assets. http://www.solarcomcapital.com/wps/wcm/connect/Solarcomcapital/Tools/TCO+Calculator/

[22] The InfraStruXure Total Cost of Ownership tool. http://www.apcc.com/tools/isx/tco/index.cfm

[23] Linux Desktop TCO Calculator. http://www.omni-ts.com/linux-desktop/tco-calculator.html.

[24] Data Center Carbon Calculator. http://www.apcmedia.com/salestools/WTOL-7DJLN9_R0_EN.swf

[25] Data Center Efficiency Calculator. http://www.apcmedia.com/salestools/WTOL-7CMGPL_R1_EN.swf

[26] Data Center Capital Cost Calculator. http://www.apcmedia.com/salestools/WTOL-7AXSAN_R1_EN.swf

[27] Virtualization Energy Cost Calculator. http://www.apcmedia.com/salestools/WTOL-7B6SFC_R0_EN.swf

[28] BROWN, D.J.A lower bound for on-line one-dimensional bin packing algorithms. Tech. Rep. No. R-864, Coordinated Sci. Lab., Univ. of Illinois, Urbana, I11., 1979.

[29] Albert Green-berg, James Hamilton, David A. Maltz, Parveen Patel. The Cost of a Cloud: Research Problems in Data Center Networks. Sigcomm 2009.

[30] The True Cost of Virtual Server Solutions. Teneja Group.