ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Improving Profitability for Cloud Service Providers Using Double Renting Scheme with Assured QOS

P.Sindhu
MTech Student
Department of CSE
Kakinada Institute of Technology and Sciences.

B.Lakshman
Assistant Professor,
Department of CSE
Kakinada Institute of Technology and Sciences.

Abstract:

In today's Computing world Cloud computing isone of the biggest innovation which uses advanced computational power and it improves data sharing and data storing capabilities. The cloud is a next generation platform that provides dynamic resource pools, virtualization, and high availability. Today, it has the ability to utilize scalable, distributed computing environments within the confines of the Internet, a practice known as cloud computing. Cloud computing is the concept implemented to decipher the daily computing problems, likes of hardware software and resource availability unhurried by computer users.

The cloud computing provides an undemanding and non-ineffectual solution for daily computing. Prevailing cloud systems mainly focus on finding an effective solution for the resource management. In cloud computing, the study of economics of the cloud is critically important. The maximization of profit is done in this. For maximizing the profit first should understand the cost and revenue. Profit maximization must consider the user satisfaction also the cost of the cloud includes the renting cost and power consumption cost. For maximizing, must reduce the cost. For this it will configure the server perfectly. For configuring the server, calculate the expected waiting time and service charge is calculated. Using this optimizing method, its demonstrated that it optimizes the speed and the size so that service providers get maximum profit.

Introduction:

Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network.

At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rack space, etc. are required for a variety of functions.

With cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications. Cloud storage is a model of data storage where the digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data. Cloud storage services may be accessed through a co-located cloud computer service, a web service application programming interface (API) or by applications that utilize the API, such as cloud desktop storage, a cloud storage gateway or Web-

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

based content management systems. Cloud storage is based on highly virtualized infrastructure and is like broader cloud computing in terms of accessible interfaces, near-instant elasticity and scalability, multi-tenancy, and metered resources. Cloud storage services can be utilized from an off-premises service or deployed on-premises. Cloud storage typically refers to a hosted object storage service, but the term has broadened to include other types of data storage that are now available as a service, like block storage. Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing. In business concepts the profit is the main factor to be exist in the field of the particular environment. Obviously, the need of profit maximization in cloud computing environment is required. 60 billion servers are currently working in this world. So the server required a huge amount of power. In order to preserve the consumption of energy, need of maximum utilization of resources is important. The important things to consider while developing such algorithm are : estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones.

This load considered can be in terms of CPU load, amount of memory used, delay or Network load. A dynamic load balancing scheme need to be proposed for avoiding over-loaded and under-loaded scenarios in cloud system. Thus the problem of dynamic Application placement should be addressed for allocating jobs to machines based on their changing demands in heterogeneous cloud environments. The profit maximization [2] is done here. In this the service charges for each service that is requested by the user is calculated. Normally between the user and server has some agreement i.e., service level agreement. In this service level agreement, defined the QoS need to provide for the user and the maximum needed execution time also. If the service provider violates this SLA no charge is provided for the particular service. So there will be the loss of the profit.

Here calculating the optimal speed and size of the input the service level agreement is provided and here a pricing model isdeveloped according to optimal size and speed and service charge is calculated. So the service provider can maximize the profit.

II. Related Work

It includes the relative mechanisms and the methods which are implemented earlier and also the advantages and disadvantages of each method is described briefly. According to the survey of the earlier mechanism, it finds that the current system implemented has more advantages. Saurabh Kumar Garg et al. Here focus on Meta scheduling of different applications from a community of users considering a commodity market. In commodity markets, service providers primarily charge the end user for services that consumes based on the value, derive from it. Pricing policies are based on the demand from the users and the supply of resources is the main driver in the competitive, commodity market models. Therefore, a user competes with other users and a resource owner with other resource owners. The financial institution Morgan Stanley is an example of a user community that has various branches across the world. Each branch has computational needs and QoS constraints that can be satisfied by Grid resources. In this scenario, it is more appealing for the company to schedule various applications in a coordinated manner.

In this work, Authors proposed two meta-scheduling online heuristics Min-Min Cost Time Tradeoff and Max-Min Cost Time Tradeoff to manage the tradeoff between overall execution time and cost and minimize them simultaneously on the basis of a tradeoff factor. The tradeoff factor indicates the priority of optimizing cost over time. These heuristics can be easily integrated in existing meta-brokers of Grid Market Infrastructures. Second, in order to study the effectiveness and efficiency of the proposed heuristics, evaluated our heuristics by an extensive simulation study. The heuristics can run in either batch mode or immediate mode. In the batch mode, the meta-broker waits for a certain time interval .

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Then at the end of the schedule interval, the meta-broker allocates all user applications to available resources. In contrast, immediate mode heuristics immediately map a task to some machine in the system for execution upon the arrival of the application. In simulation, studied the heuristics in batch mode. For scheduling approaches outside Grid computing, Min-Min, Min-Max and Suffrage are three major task-level heuristics employed for resource allocation. As developed based on specific domain knowledge, cannot be applied directly to Grid scheduling problems, and hence have to be enhanced accordingly. The main contribution are thus to design two heuristics to manage and optimize the tradeoff between cost and execution time of user application in a concurrent user's environment for Utility Grids. Adopt some ideas from MinMin and Min-Max heuristics to design my algorithm. Qian Zhu et al. While current cloud systems are beginning to offer the utility-like provisioning of services, provisioning of resources has to be controlled by the end users. It is desirable that resource allocation in a cloud environment can be performed automatically and dynamically, based on users' high-level needs.

The allocation of resource to each VM can be dynamically controlled, and the resource costs incurred depend upon the resources allocated. Furthermore, aresource model is proposed to map any given combination of values of adaptive parameters to resource requirements in order to guarantee that the resource cost stays under the budget. Gemma Reig et al. Here a prediction system to determine the minimum job resource requirements to be executed before its deadline. One key innovation of the prediction system is the usage of Machine Learning to enable the translation from service-level metrics to resource requirements. Enabling the cloud to non-expert IT users by means of using service-level metrics and help providers to do a smart utilization of their resources by using the resources left by web applications to execute jobs in an efficient way e.g. discard jobs in advance,avoiding the risk of wasting resources in executing jobs that will not meet their deadlines.

The Scheduler accepts incoming jobs and web applications to be planned. It queries the Prediction System and it decides, depending on the policy being used and the resources status, how to allocate resources to the incoming jobs and how to elastically size up and down the resource allocation for web applications, in order to fulfill their respective QoS. Ana Maria Oprescu et al. BaTS,budget, constrained scheduler. BaTS can schedule large bags of tasks onto multiple clouds with different CPU performance and cost. BaTS schedules such that a bag of tasks will be executed within a given budget, while minimizing the completion time.BaTS requires no a-priori information about task completion times, instead BaTS learns application throughput at run time, using an initial sampling phase and a moving average throughout the computation. BaTS are scheduling large bags of tasks onto multiple cloud platforms. The core functionality is to allocate a number of machines from different clouds, and to adapt the allocation regularly by acquiring or releasing machines in order to minimize the overall makespan while respecting the given budget limitation machines. Assume that the tasks of a bag are independent of each other, so they are ready to be scheduled immediately. Also assume,The individual tasks are scheduling a round-robin manner onto the allocated machines. Assume that the tasks can be preempted and rescheduled later, if needed by a reconfiguration of the cloud environment. Task model incurs no prior knowledge about the task execution times.

Existing System:

In Many existing research they only consider the power consumption cost. As a major difference between their models and ours, theresource rental cost is considered in this paper as well, since it is a major part which affects the profit of service providers.The traditional single resource renting scheme cannot guarantee the quality of all requests but wastes a great amount of resources due to the uncertainty of system workload. To overcome the weakness, we propose a double renting scheme as follows, which not only can

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

guarantee the quality of service completely but also can reduce the resource waste greatly.

Proposed System:

In this section, we first propose the Double-Quality-Guaranteed (DQG) resource renting scheme which combines long-term renting with short-term renting. The main computing capacity is provided by the long-term rented servers due to their low price. The short-term rented servers provide the extra capacity in peak period

Advantages:

In proposed system we are using the Double-Quality-Guaranteed (DQG) renting scheme can achieve more profit than the compared Single-Quality-Unguaranteed (SQU) renting scheme in the premise of guaranteeing the service quality completely.

Problem Statement:

A profit maximization function is defined to find an optimal combination of the server size R and the queue capacity K such that the profit is maximized. However, this strategy has further implications other than just losing the revenue from some services, because it also implies loss of reputation and therefore loss of future customers. In , Cao et al. treated a cloud service platform as an M/M/m model, and the problem of optimal multiserver configuration for profit maximization was formulated and solved. This work is the most relevant work to ours, but it adopts a single renting scheme to configure a multiserver system, which cannot adapt to the varying market demand and leads to low service quality and great resource waste. To overcome this weakness, another resource management strategy is used in , which is cloud federation. Using federation, different providers running services that have complementary resource requirements over time can mutually collaborate to share their respective resources in order to fulfill each one's demand . However, providers should make an intelligent decision about utilization of the federation (either as a contributor or as a consumer of resources) depending on different conditions that they might face, which is a complicated problem.

Scope:

In this paper, we only consider the profit maximization problem in a homogeneous cloud environment, because the analysis of a heterogenous environment is much more complicated than that of a homogenous environment. However, we will extend our study to a heterogenous environment in the future.

Architecture:

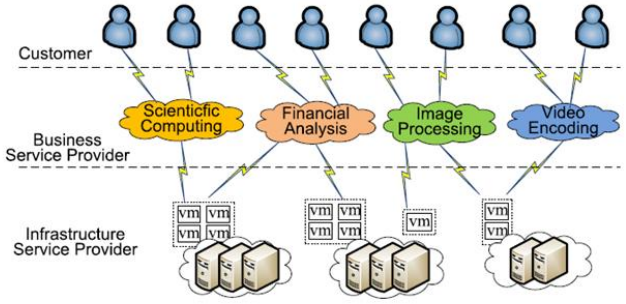

Fig. 1: The three-tier cloud structure.

Implementation of Modules:

1. Cloud computing,
2. queuing model.
3. Business Service Module
4. Cloud customer Module.
5. Infrastructure Service Provider Module.

Cloud Computing:

Cloud computing describes a type of outsourcing of computer services, similar to the way in which the supply of electricity is outsourced. Users can simply use it. They do not need to worry where the electricity is from, how it is made, or transported. Every month, they pay for what they consumed. The idea behind cloud computing is similar: The user can simply use storage, computing power, or specially crafted development environments, without having to worry how these work internally. Cloud computing is usually Internet-based computing. The cloud is a metaphor for the Internet based on how the internet is described in

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

computer network diagrams; which means it is an abstraction hiding the complex infrastructure of the internet. It is a style of computing in which IT-related capabilities are provided "as a service", allowing users to access technology-enabled services from the Internet ("in the cloud")without knowledge of, or control over the technologies behind these servers.

Queuing model:

we consider the cloud service platform as a multiserver system with a service request queue.The clouds provide resources for jobs in the form of virtual machine (VM). In addition, the users submit their jobs to the cloud in which a job queuing system such as SGE, PBS, or Condor is used. All jobs are scheduled by the job scheduler and assigned to different VMs in a centralized way. Hence, we can consider it as a service request queue. For example, Condor is a specialized workload management system for computeintensive jobs and it provides a job queueing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management. Users submit their jobs to Condor, and Condor places them into a queue, chooses when and where to run them based upon a policy.An M/M/m+Dqueueing model is build for our multiserver system with varying system size. And then, an optimal configuration problem of profit maximization is formulated in which many factors are taken into considerations, such as the market demand, the workload of requests, the server-level agreement, the rental cost of servers, the cost of energy consumption, and so forth. The optimal solutions are solved for two different situations, which are the ideal optimal solutions and the actual optimal solutions.

Business Service ProvidersModule:

Service providers pay infrastructure providers for renting their physical resources, and charge customers for processing their service requests, which generates cost and revenue, respectively. The profit is generated from the gap between the revenue and the cost.Inthis module the service providers considered as cloud brokers because they can play an important role in between cloud customers and infrastructure providers ,and he can establish an indirect connection between cloud customer and infrastructure providers.

Infrastructure Service Provider Module:

In the three-tier structure, an infrastructure provider the basic hardware and software facilities. A service provider rents resources from infrastructure providers and prepares,a set of services in the form of virtual machine (VM). Infrastructure providers provide two kinds of resource renting schemes, e.g., long-term renting and short-term renting. In general, the rental price of long-term renting is much cheaper than that of short-term renting.

Cloud Customers:

A customer submits a service request to a service provider which delivers services on demand. The customer receives the desired result from the service provider with certain service-level agreement, and pays for the service based on the amount of the service and the service quality.

Conclusion:

Maximize the profit of service providers, this paper has proposed a novel Double-Quality-Guaranteed (DQG) renting scheme for service providers. This scheme combines short-term renting with long-term renting, which can reduce the resource waste greatly and adapt to the dynamical demand of computing capacity. An M/M/m+Dqueueing model is build for our multiserver system with varying system size. And then, an optimal configuration problem of profit maximization is formulated in which many factors are taken into considerations, such as the market demand, the workload of requests, the server-level agreement, the rental cost of servers, the cost of energy consumption, and so forth. The optimal solutions are solved for two different situations, which are the ideal optimal solutions and the actual optimal solutions. In addition, a series of calculations are conducted to compare the profit obtained by the DQG renting scheme with the Single-Quality-Unguaranteed (SQU)

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

renting scheme. The results show that our scheme outperforms the SQU scheme in terms of both of service quality and profit.

References:

[1] Jing Mei, Kenli Li, Member, IEEE, AijiaOuyang and Keqin Li, Fellow, IEEE, A Profit Maximization Scheme with Guaranteed Quality of Service in Cloud Computing, IEEE Transactions on Computers (Volume:64 , Issue: 11)

[2] [2] V.Ashok Gajapathi Raju & G.Nagendra Kumar, An Assessment on Cloud Computing Data Security and Exploration, IJMETMR, Volume No: 2 (2015), Issue No: 5 (May), http://www.ijmetmr.com/olmay2015/VAshokGajapathiRaju-GNagendraKumar-61.pdf

[3]Qian Zhu, Student Member, IEEE, and Gagan Agrawal, Senior Member, IEEE, "Resource Provisioning with Budget Constraints for Adaptive Applications in Cloud Environments," ieee transactions on services computing, vol. 5, no. 4, october-december 2012.

[4] Pankesh Patel, Ajith Ranabahu AmitSheth, "Service Level Agreement in Cloud Computing," Knoesis Center, Wright State University, USA.

[5] Saurabh Kumar Garg, Raj kumar Buyya1 and H. J. Siegel " Scheduling Parallel Application On Utility Grids: Time And Cost Trade-off management" Management" The University of Melboune Victoria 3010, Australia.

[6] Gemma Reig, Javier Alonso, and Jordi Guitart "Prediction Of Job Resource Requirements For Deadline Schedulers to Manage High-Level SLAs On The Cloud", 2010 Ninth IEEE International Symposium on Network Computing and Applications.

[7] J. Mei, K. Li, J. Hu, S. Yin, and E. H.-M. Sha, "Energy- aware preemptive scheduling algorithm for sporadic tasks on dvs platform," MICROPROCESS MICROSY., vol. 37, no. 1, pp. 99–112, 2013.

[8] S. Liu, S. Ren, G. Quan, M. Zhao, and S. Ren, "Profit aware load balancing for distributed cloud data centers," in IEEE 27th Int'l Symp. Parallel & Distributed Processing. IEEE, 2013, pp. 611–622.

[9] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design," IEICE Trans. Electronics, vol. 75, no.4, pp. 371–382, 1992.

[10] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," in Proc. 41st annual Design Automation Conf. ACM, 2004,pp.