International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
International Conference on Industrial Automation and Computing (ICIAC- 12-13th April 2014)

RESEARCH ARTICLE OPEN ACCESS

A Survey on Performance Analysis for Optimization of Storage Reallocation Strategies in Cloud Environment

Pallavi Shelke*, Prof. M.B.Gudadhe **

*Department of Computer Science & Engineering, S.V.P.C.E.T College, RTMNU University, Nagpur,Maharashtra.India
pallavishelke12@gmail.com
** Department of Computer Science&Engineering,S.V.P.C.E.T College,RTMNU University, Nagpur,Maharashtra ,India
mbgpatil@yahoo.com

ABSTRACT
In Cloud Computing environment, resource allocation strategies are used to improve overall system utilization and data replication is used to improve the data access This environment is used by most of IT industries to meet their high computational requirement with reduce cost. To achieve this, efficient resource allocation strategies are needed. This paper gives the development of resources allocation strategies which mainly focus on three auction process strategies. Advanced reservation, Overbooking, Just In Time bidding which is under the pre auction and during auction process. This paper consider Stroage as resources and dynamic replication strategy to store the cloud user data ,Paper would also work on read and write to given storage with minimum delay.
Keywords -cloud computing, Economic, Resource Allocation, Storage, High utilization strategies

I. INTRODUCTION

Cloud computing, where application and files are hosted on"Cloud"consisting of thousands of computer and servers, all linked together and accessible via internet. .On cloud infinite computing resources available that may used on demand..User or customer can access the stored data at any time by using application programming Interface (API) provided by cloud provider through any terminal equipment connected to the internet. Resources are nothing but Storage, Hardware, Software that are available as services. These Services provided by service provider which classified into three model Infrastructure as service , Platform as service and Software as service .

Computational economies have long been advertized as a means of allocating resources in both centralized and decentralized computing system [4].Such system many times includes poor performance,high latency and high overhead example ,in competitive economy resources are typically "reserved by m participants for the duration of a negotiation. In most cases ,there are only a n "winning" participants, therefore all other "m-n" reservation are essentially wasted for the duration of that negotiation. Moreover, there is an opportunity cost to reserving resources during a negotiation., as they will not be available for other negotiations that begins during the interval of the first negotiation. This type of scenario is clearly evident in auction or tender markets, however it can also seen in any negotiation in which parties are competing against one another for the goods on offer. In any case, this wasteful negotiation process is expensive in both time and cost and therefore reduces the overall utilization of System.

In this paper, these inefficiencies are addressed by two general principles:

1) Avoid the commitment of resources
2) Avoid repeating negotiation and allocation process.

So these inefficiencies overcome by using High performance resource Utilization strategies, like overbooking, advanced reservation, just in time bidding, all are come under pre auction and during auction process.This paper looks at these strategies as a means to improve the performance of auction system in allocation of resources in high performance computing Context[1].

The rest of paper is organized as follows: In Section II Various resources allocation strategies and their impacts in cloud environment are discussed . In section III Related work is discussed and IV is conclusion .

II. LITERATURE SERVEY

Cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers.

Significance of resources Allocation cloud computing.

In cloud allows business customers to scale up and down their resource usage based on needs. So resources allocation (RA) is the process of assigning available resources to the needed cloud application over the Internet .Resource allocation starves services if the allocation not managed precisely. So different resource allocation strategies are proposed by different author [3]. Resources allocation integrates cloud provider activities for utilizing and allocating scarce resources with limit of of cloud application. Resource allocation requires amount and type of resources for different type of application of in order to complete a user job. Criteria used for testing resource allocation Access,Fault tolerance,.Power distribution reliability, scalability resource contention,Scarcity,resource fragmentation ,over provisioning, under provisioning .

Resource Allocation Strategies (Ras) At A Glance[2]

The input parameter to RAS and the way of resource allocation vary based on services, infrastructure and nature of application which demand resources.[3]This paper summarizes the classification of RAS and its impacts in Cloud Environment.

Following are types Resource Allocation Strategies: Execution time ,Policy, VM(virtual machine),Gossip,Utility Hardware Resource Dependency ,Auction Application ,SLA(service level agreement).

A. Execution Time

Different kinds of resource allocation mechanisms are proposed in cloud. Actual task execution time and preempt able scheduling is considered for resource allocation[7]. It overcomes the problem of resource contention and increases resource utilization by using different modes of renting computing capacities. But the VM model considered in [7] is heterogeneous and proposed for IaaS.

B. Policy

Since centralized user and resource management lacks in scalable management of users, resources and organization- level security policy. Paper [8] has proposed a decentralized user and virtualized resource management for IaaS by adding a new layer called domain in between the user and the vritualized resources. Based on role based access control(RBAC), virtualized resources are allocated to users through domain layer [9], which used the most-fit processor policy for resource allocation. The most-fit policy allocates a job to the cluster, which produces a leftover processor distribution, leading to the most number of immediate subsequent job allocations. Experimental results shows that the most-fit policy has higher time complexities but the time overheads are negligible compared to the system long time operation. This policy is practical to use in a real system.

C. Virtual Machine (VM)

A system which can automatically scale its infrastructure resources is designed in [10]. The system composed of a virtual network of virtual machines capable of live migration across multi-domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources. But the above work considers only the non- preemptable scheduling policy. Allocation of the resources based on the speed and cost of different VMs in IaaS is discuss in paper [11] which allowing the user to select VMs and reduces cost for the user.

D. Gossip

Cloud environment differs in terms of clusters, servers, nodes, their locality reference and capacity. The problem of resource management for a large-scale cloud environment (ranging to above 100,000 servers) is addressed in [12] and general Gossip protocol is proposed for fair allocation of CPU resources to clients .In paper [13], Gossip based co-operative VM management with VM allocation and cost management is introduced. By this method, the organizations can cooperate to share the available resources to reduce the cost. Here the cloud environments of public and private clouds are considered. They have formulated an optimization model to obtain the optimal virtual machine allocation. A gossip-based protocol for resource allocation in large- scale cloud environments is proposed in [14]. It performs a key function within distributed middle ware architecture for large clouds.

E. Utility Function

There are many proposals that dynamically manage VMs in IaaS by optimizing some objective function such as minimizing cost function, cost performance function and meeting QoS objectives. The objective function is defined as Utility property which is selected based on measures of response time, number of QoS, targets met and profit etc.

There are few works [15],[16] that dynamically allocate CPU resources to meet QoS objectives by first allocating requests to high priority applications. For multitier Cloud computing systems

(heterogeneous servers), resource allocation based on response time as a measure of utility function is proposed by considering CPU, memory and communication resources in paper [17].

F. Hardware Resource Dependency

In paper [18], to improve the hardware utilization, Multiple Job Optimization (MJO) scheduler is proposed. Jobs could be classified by hardware-resource dependency such as CPU- bound, Network I/O-bound, Disk I/O bound and memory bound. MJO scheduler can detect the type of jobs and parallel jobs of different categories. Based on the categories, resources are allocated. This system focuses only on CPU and I/O resource.

G. Auction

Cloud resource allocation by auction mechanism is addressed in [19]. The proposed mechanism is based on sealed-bid auction. The cloud service provider collects all the users' bids and determines the price. The resource is distributed to the first kth highest bidders under the price of the (k+1)th highest bid. This system simplifies the cloud service provider decision rule and the clear cut allocation rule by reducing the resource problem into ordering problem. But this mechanism does not ensure profit maximization due to its truth telling property under constraints. The aim of resource allocation strategy is to maximize the profits of both the customer agent and the resource agent in a large data center by balancing the demand and supply in the market. It is achieved by using market based resource allocation strategy in which equilibrium theory is introduced (RSA-M) [20]. RSA-M determines the number of fractions used by one VM and can be adjusted dynamically according to he varied resource requirement of the workloads. One type of resource is delegated to publish the resource's price by resource agent and the resource delegated by the customer agent participates in the market system to obtain the maximum benefit for the consumer. Market Economy Mechanism is responsible for balancing the resource supply and demand in the market system.

H. Application

Resource Allocation strategies are proposed based on the nature of the applications in [21] [22]. In the work by Tram et al. [33], Virtual infrastructure allocation strategies are designed for work flow based applications for this work flow based applications, the application logic can be interpreted and exploited to produce an execution schedule estimate. This helps the user to estimate the exact amount of resources that will be consumed for each run of the application. Four strategies such as Naive, FIFO, Optimized and services group optimization are designed to allocate resources and schedule computing task

Database replicas allocation strategy is designed in [23]. In that work, the resource allocation module divides the resource (CPU, Memory and DB replicas) allocation problem in two levels. The first level optimally splits the resources among the clients whereas the database replicas are expandable (dynamic) in the second level, based on the learned predictive model. It achieves optimal resource allocation in a dynamic and intelligent fashion.

I. SLA

In cloud, the works related to the SaaS providers considering SLA are still in early stage of Existence. Therefore in order to achieve the SaaS providers' objective, various RAS specific to SaaS in cloud has been proposed. With the emergence of SaaS, applications have started moving away from pc based to web delivered-hosted services. Most of the RAS for SaaS focused towards customer benefits. [24] have mainly considered QoS parameters on the resource provider's side such as price and offered load.

III. RELETED WORK

Auction System is powerful tool to resolve a competitive situation of resource requirement. This paper mainly focus on market based auction process .This work concentrate on high performance resource utilization strategies and that can be deployed in cloud , mainly Overbooking,Advanced reservation and Just in time bidding. Each of the strategies is examined experimentally with in the context of market based cloud .

I OVERBOOKING [1][2][5]

Overbooking strategy used in Economic model can improved the system utilization rate and occupancy..This strategy is capable to allocate the resources dynamically so easily overcome the problem of resource allocation .so this strategy used by in auction based resource allocation model for a resource provider .Paper [25] overbooking model from revenue Management to manege cancellation and no-show of reservation in a grid system. So Same technique can be used for cloud environment for resource allocation. Many time resource provider is faced with a prospect of loss of income and lower system utilization because during period of high demands resource provider reject booking from potential user,who are committing to use the resources and willing to pay for higher price..so overbooking offers solution for above problem ,by allowing the resource provider to accepts more reservation than capacity .Hence it can be effectively used to minimize the loss of income.[26,27].In a

study [28]on American Airline ,50% of the booking were result in cancellation or no-shows. many time it found that 15% of the flight seats would be unused, if booking were only limited to the capacity of plane. Most airlines routinely overbook aircraft in an attempt to maximize occupancy and therefore revenue by ensuring they have maximum no of passengers on a flight. without overbooking full flight often depart with 15% of seats empty[29].In Paper [25] various overbooking model models explained ,also its integrates overbooking into a capacity allocation heuristic to estimate a suitable quota for for different user ,also calculate cost and penalties of each reservation. Proposed research work take a help of paper[25] to implement overbooking algorithm also perform different performance analysis.

Based on different survey and study this paper focus on,Overbooking Strategy is to improved the expected profit,With overbooking resources /services are get on time.(i.e hundred percent service to cloud users), Waiting period of services is less , Most of the time resources are wasted,Need to calculate workload every time.

II. ADVANCED RESERVATION [1][2][5]

Advanced reservation strategy important to guarantee that resources are available at the expected time. This strategy support in distributed resource allocation environment for performance predictability, meeting re-source requirements, and providing Quality of Service (QoS) guarantees [30,31,33,34,11].Paper [31] propose Globus Architecture for Reservation and Allocation(GARA) to address various issues like dynamic discovery ,advance or immediate reservation of resources, that will often be heterogeneous in type and implementation an controlled and administered. Paper[34] introduces a real options approach to model grid markets and how to handle grid resource reservation in a reasonable economic fashion.Paper[12]presents a method to allocate resources for real-time tasks using the "Infrastructure as a Service" model.

Based on different survey and study this paper focus on Advanced reservation improved the expected profit, Resources are not wasted,With Advanced reservation resources are get on time,waiting period of getting resources/service is less .

III. JUST IN TIME BIDDING[1][5]

Just-In-Time (JIT) bidding (or "sniping") is employed to minimize the effect of auction latency by committing/reserving resources to be allocated by the resource provider at the last possible moment. Just in time bidding is used during the auction process. It may happen that during the negotiation resource state may change and therefore invalidate (cancel) provider valuation (or bid). In general there are two way to ways to minimize the effect of latency .

- Reducing the duration of auction. The problem with this approach is that there is minimal time for providers to discover the auction and to compute their bids.
- Bid as late as possible the advantage with this approach is that provider can compute their bids with the most up to date resources state and resources are reserved for shorter time .Primary problem with this approach is time sensitivity, the auction can missed if the bid is too late or experiences unexpected network delays.[1]

For just in time bidding process we consider the point Profit maximization due to truth telling property, under constrains ,Less overhead, Based on agreement forty percent time resources and services are available for cloud users .

IV. CONCLUSION

This paper has summaries the different resource allocation strategies and its impact on cloud environment. Paper also focused on three main strategies overbooking, advanced reservation ,just in time bidding which used under the auction process, In future work this paper would proposed the hybrid solution by combing these three main strategies and used it for storage reallocation in cloud environment. By giving such type of solution we just try to improved maximum utilization of system, improves cloud provider benefits ,as well as fast access of storage resources to each cloud user within short time.

REFERENCES

[1]. Kyle Chard,Kris Bubendorfer,"High Performanceresource allocation strategies for computational Economies ",IEEE Transaction on parallel and distributed system,VOL 24,No 1.,January 2013.

[2]. Jeyaram.G,Vidhya V"Efficient Resource Allocation Stratergies for cloud Data Centers "IJARCCE,Vol 2,Issue 2,February 2013 .

[3]. V.Vinpthina ,Dr. R. Sridardan,Dr. Padmavathi ganpathi, "A servey on Resource Allocation strategies cloud computing" IJACSA ,vol .3,No 6.2012

[4]. I.E. Sutherland"A futures Market in Computer Time",Comm. ACM ,vol..11,No 6,pp. 449-451,1968.

[5]. Kyle Chard,Kris Bubendorfer,Peter Komisarczuk,,"High Occupancy Resources

allocation for Grid and cloud system, a study with Drive . HPDC' '10 ,June 20-25,2010,chicago,illinois,USA. ACM 978-1-60558- 942-8/10/06.

[6]. G. Birkenheuer, A. Brinkmann, and H. Karl. The gain of overbooking. In Lecture Notes in Computer Science(LNCS), Job Scheduling Strategies for Parallel Processing, volume 5798/2009, pages 80–100, 2009.

[7]. Jiyani et al.: Adaptive resource allocation for preemptable jobs in cloud systems (IEEE, 2010), pp.31-36.

[8]. Dongwan Shin and HakanAkkan :Domain- based virtualized resource management in cloud computing

[9]. Kuo-Chan Huang &Kuan-Po Lai: Processor Allocation policies for Reducing Resource fragmentation in Multi cluster Grid and Cloud Environments (IEEE, 2010), pp.971-976.

[10]. P.Ruth,J.Rhee, D.Xu, R.Kennell and S.Goasguen, "Autonomic Adaptation of virtual computational environments in a multi-domain infrastructure", IEEE International conference on Autonomic Computing, 2006,pp.5-14

[11]. Karthik Kumar et al.: Resource Allocation for real time tasks using cloud computing (IEEE, 2011), pp.

[12]. RerngvitYanggratoke, FetahiWuhib and Rolf Stadler: Gossip-based resource allocation for green computing in Large Clouds: 7th International conference on network and service management, Paris, France, 24-28 October, 2011.

[13]. DusitNiyato, Zhu Kun and Ping Wang : Cooperative Virtual Machine Management for Multi-Organization Cloud Computing Environment

[14]. FetahiWuhib and Rolf Stadler : Distributed monitoring and resource management for Large cloud environments(IEEE,2011),pp.970-975.

[15]. D. Gmach, J.RoliaandL.cherkasova, Satisfying service level objectives in a self-managing resource pool. In Proc. Third IEEE international conference on self-adaptive and self organizing system.(SASO'09) pages 243-253.IEEE Press 2009

[16]. X.Zhu et al. Integrated capacity and workload management for the next generation data center. In proc.5th international conference on Automatic computing(ICAC'08),pages 172-181,IEEE Press 2008

[17]. Minarolli and Bernd Freisleben: Uitlity – based Resource Allocations for virtual machines in cloud computing(IEEE,2011),pp.410-417

[18]. Weisong Hu et al. : Multiple Job Optimization in MapReduce for Heterogeneous Workloads : IEEE Sixth International Conference on Semantics, Knowledge and Grids 2010,pp.135-140.

[19]. Wei-Yu Lin et al. : Dynamic Auction Mechanism for Cloud Resource Allocation: 2010 IEEE/ACM 10th International Conference on Cluster, Cloud and Grid Computing, pp.591-592

[20]. Xindong YOU, Xianghua XU, Jian Wan, Dongjin YU:RAS-M :Resource Allocation Strategy based on Market Mechanism in Cloud Computing(IEEE,2009),pp.256-263.

[21]. Tram Truong Huu& John Montagnat: Virtual Resource Allocations distribution on a cloud infrastructure (IEEE, 2010), pp.612-617

[22]. WaheedIqbal, Matthew N.Dailey, Imran Ali and Paul Janecek& David Carrera: Adaptive Resource Allocation for Back-end Mashup Applications on a heterogeneous private cloud

[23]. PenchengXiong,Yun Chi, Shenghuo Zhu, Hyun Jin Moon, CaltonPu& Hakan Hacigumus: Intelligent Management Of Virtualized Resources for Database Systems in Cloud Environment(IEEE,2011),pp.87-98.

[24]. I.Popovici et al,"Profitable services in an uncertain world". In proceedings of the conference on supercomputing ,CSC,2005

[25]. A. Sulistio, K. H. Kim, and R. Buyya. Managing cancellations and no-shows of reservations with overbooking to increase resource revenue. In CC GRID'08: Proceedings of the Eighth IEEE Internationa Symposium on Cluster Computing and the Grid,pages267,276,Was ington, DC, USA, 2008. IEEE

[26]. J. I. McGill and G. J. V. Ryzin. Revenue Management,Research Overview and Prospects. *Transportation Science* 33(2):233.256, 1999.

[27]. R. L. Phillips. *Pricing and Revenue Optimization*. Stanford University Press, 2005

[28]. B. C. Smith, J. F. Leimkuhler, and R. M. Darrow Yield Management at American Airlines. *Interfaces*, 22:8.31, 1992

[29]. B. C. Smith, J. F. Leimkuhler, and R. M. Darrow.Yield Management at American Airlines.INTERFACES, 22(1):8{31, 1992

[30]. C. Castillo, G. Rouskas, and K. Harfoush. Efficient resource management using advance reservations for heterogeneous grids. In IPDPS'08: Proceedings of the IEEE International Symposium on Parallel And Distributed Processing, pages 1-12, April 2008

[31]. I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, and A. Roy. A distributed resource management architecture that supports advance reservations and co-allocation. In

IWQoS'99:Proceeings of the Seventh International Workshop on Quality of Service, 1999.

[32]. M. Siddiqui, A. Villaz_on, and T. Fahringer. Grid capacity planning with negotiation-based advance reservation for optimized qos. In SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing, page 103, New York, NY, USA, 2006. ACM.

[33]. A. S. Mcgough, A. Afzal, J. Darlington, N. Furmento A.Mayer, and L. Young. Making the grid predictabl through reservations and performance modelling. The Computer Journal, 48(3):358{368, 2005.

[34]. T. Meinl. Advance reservation of grid resources via Real options. In Proceedings of the 10th IEEE Conference on E-Commerce Technology and the 5thIEEE Conference on Enterprise Computing, E-Commerce and E-Services, pages 3{10, July 2008.