Survey on Resource Allocation Techniques for Query Processing in the Cloud Computing

Sampat peddi[#1] andDr.Bharati Harsoor[*2]
[#]*Dept.of Information Science and Engg, PDA College of Engg. Kalaburagi, India,*
[*]prof & Head, Dept. of Information Science and Engg, PDA College of Engg. kalaburagi,India.

Abstract–Cloud computing is one of the fastest emerging Technologyamong IT user and it provides a large verity of resource to its customers by cloud providers, the resource is allocated on demand to the customers. it is difficult to the cloud providers as it has a limited or finite number of resources. In this paper we present several RA techniques which concentrate on the key issuesrelated to existing resource allocation in the cloud and the discussion about resource allocation techniques for cloud user in the cloud and benefit for the researcher to explore more about the resource allocation in the cloud environment

Key words: cloud computing, resource allocation.

I. INTRODUCTION - Today a developing number of organizations need to process in huge amount of information in a cost-effective way. Great agents for these organizations are administrators of Internet web crawlers, similar to Google, Yahoo, or Microsoft. The tremendous measure of information they need to manage each day has made customary database arrangements restrictively costly. Rather, these organizations have advanced a design worldview dependent on countless product servers. Issues like preparing crept reports or recovering a web file are part into a few free subtasks, disseminated among the accessible hubs, and registered in parallel. So as to improve the advancement of disseminated applications over such structures, a considerable lot of these organizations have additionally fabricated modified information preparing systems. Models are Google's Map Reduce , Microsoft's Dryad , or Yahoo's! Map-Reduce-Merge.

They can be grouped by terms like high throughput Computing (HTC) or many-task computing (MTC), contingent upon the measure of information and the quantity of errands associated with the calculation. Despite the fact that these frameworks contrast in plan, their programming models share comparable targets, in particular concealing the problem of parallel programming, adaptation to internal failure, and execution enhancements from the designer. Designers can commonly keep on composing successive projects. The preparing structure at that point deals with circulating the program among the accessible hubs and executes each example of the program on the proper part of information. For organizations that just need to process a lot of information once in a while running their very own server farm is clearly impossible. Rather, Cloud registering has risen as a promising way to deal with lease an enormous IT framework on a momentary compensation for every utilization premise. Administrators of purported Infrastructure-as-a-Service (IaaS) mists, similar to Amazon EC2 , let their clients dispense, access, and control a lot of virtual machines (VMs) which keep running inside their server farms and just charge them for the timeframe the machines are assigned.

II. RESOURCE ALLOCATION- Resource allocation in cloud is the process of assigning cloud resources to the requested cloud users over the internet. Cloud computing is mainly popular for flexibility which can be attained through scalability and availability of services whenever in need. It

aims to supply reliable, customized and good QoS parameters in the dynamically changing computing environments. To attain these goals, resource allocation is very much important. Proper resource allocation gains customer satisfaction

Challenges in resource allocation

- Resource contention: It mainly arises when the two or more cloud enabled applications are requesting resources to the service provider.
- Resource scarcity: It mainly arises when they are limited number of resources that are available in the cloud service Provide
- Resource fragmentation: This situation mainly arises whenever the cloud application requests resources to the server and the resources are isolated

Which means there will be enough resources but not able to allocate to the Needed application

- Over provisioning: Cloud service consumers get more resources from the service Providers than the demanded resources
- Under provisioning: It mainly occurs when the application is assigned to less number of resources than the demand.

Table 2.1.Comparison of the existing Resource Allocation Techniques in Cloud Computing

Technique Used	Parameters Used	Advantages	Future Work
Tenant-based	Number of VM instances; Workload pikes (increment and peak based); VM capacity	Average Underutilized are statistically improved	Can be performed on other different scenarios such as HPC
Loyaltybased	Transaction rate	Improve the successful transaction rate of the system under the environment of cloud computing	Can be performed taking other scenarios
Agent-based	Average number of visiting locations for a request; average geographical distance after allocation; average allocation time for a request; success rate; average number of request denial for success	It has high performance in terms of average allocation time and geographical distance. The model has fast allocation time	Work can be done on different size of memory spaces, live migration and dynamic workloads
Economy based	Profit, amount of accepted lease, execution cost of the lease	Reduces the cost of execution of the consumer's lease and increases the profit of the provider to a considerable extent	Can be performed considering other scenarios
Rule-based	Access resources on demand, Resource utilization of	Effective increase in resource	This technique can be used on other services in

	instances (VM in AWS) Virtual machines(Microsoft Azure)	utilization and implemented in IaaS ,increase the scalability of private cloud on-demand and decrease the cost in hybrid environment	Iaas environment
Utility-based	Demand for web server in VM; number of active nodes; VM web server response time; global system utility	Effective cost reduction while improving global system utility using VM migration with better performance	Reliability cost can be included and inclusion of memory, disk and network in distributed environment
Locationbased	Response time, Transfer Time, Processing Time	Reduction of the latency with reducing cost and high user satisfaction	Including the designed control entity for adding and releasing instances in the cloud
Topologybased	Job completion time; efficiency	Reduce Job completion time up to 59% then simple techniques	Increase the objective functions by involving infrastructure costs, power and reliability
Gossip-based	Scalability	Minimizing power consumption through server consolidation when the system is in under load and fair resource allocation in case of overload	Determination of the convergence rate of GRMP-Q and its dependence on CPU and memory demands and to implement in heterogeneous environment making protocol robust from machine failures.
Preferencebased	Actual payment of a winner with his bestprice payment, variation in utilities earned by winners, Revenues Earned	High Performance benefit in revenues to service provider and payments of cloud users besides ensuring an optimum resources use.	Including an energy efficient scheduling strategy to allot auction winners' tasks to suitable VMs

Related work

There are a number of works for analyzing resource allocation in cloud platforms. Allocations of resources based on various techniques have been attempted. Resources are allocated in cloud considering numerous parameters such as high throughput, maximum efficiency, maximum energy and power consumption etc. In the following, a quick review of some of the works that are directly related to resource allocation is discussed.

Tenant-based

This technique is based on a resource allocation model to arrange SaaS applications over cloud computing platforms and to create a cost-effective scalable
environment by taking into account their multi-tenancy. In this technique, to establish formal measurements for over and under provisioning of virtualized resources in cloud infrastructures, specifically for SaaS platform
deployments, three complementary approaches are used:
(1) tenant-based load balancing which distributes requests according to the tenant information, (2) tenant-based isolation which encapsulates the execution of each tenant, and (3) a tenant-based VM instance allocation which determines the number of VM instances needed for certain workload which is based on VM capacity and tenant context weight.

Loyalty- based

This technique uses Master-Slave framework which is a role based access control considering the trust of the node and meets the requirements using the services. This technique provides assessment according to the real-time condition of the system and then allocates resources based on their assessment generated. This kind of dynamic feedback mechanism guarantees the firmness of the system and reliability of the services

Agent-based

This technique is based on adaptive resource allocation model in which resources are allocated based on two evaluations: 1) the workload of data centers, and 2) the
geographical distance between a consumer and a data center. Agent based test bed is used to allocate VM to the closet data center to the location of consumers. For this reason, the service provider may also guarantee the fast response time in future VM execution .

Economy-based

In this technique, an economy based leasing algorithm is developed which aims at focusing on incentives of both the providers and the consumers. This economy customer and service provider using the concept of optimization techniques to optimize the costs.

Rule-based This technique is based on IaaS layer to access resources on demand. Rule Based Resource Manager is developed which increase the scalability of private cloud on-demand and decrease the cost in hybrid environment and also set the time for private cloud and public cloud so as to fulfill the request and provide the services in time.

Utility-based

This technique is based on a two-tier resource manager which is based on local node utility functions and a global system utility function that considers VM livemigration as an important part of resource allocation mechanism. The resource manager manages an IaaS Cloud environment comprising of several VMs runningon top of physical machines. Utility Function model is developed in this technique and dynamically allocation of CPU resources to VMs in IaaS Clouds taking into account QoS objectives and operating costs .

Location-based

In this technique, the data of social Web application Twitter is used to achieve a load forecast on a particular service so that the significant probability for load isidentified. By this, the main demand at a location can be easily mapped to an optimal hosting region in such a way that unnecessary latencies can possibly omitted because of shorter routing distance and resources can be instantiated at the main region of load. This technique provides an optimized trade-off between high user satisfaction and low emerging costs. That will lead to anoptimal covering of resource demands as well as a decrease of the latency during service provisioning .

Topology-based

In this technique, a prototype for Topology-Aware Resource Allocation (TARA) is developed and evaluated it on a 80 server cluster with two representativeMap Reduce-based benchmarks. This technique is based on adopts a "what if" methodology so as to guide allocation decisions taken by the IaaS. The architectureuses a prediction engine with a lightweight simulator to estimate the performance of a genetic algorithm given and a resource allocation to find an enhanced solution in the large search space.

Gossip-based

This technique is based on a generic gossip protocol GRMP-Q for resource allocation, which can be instantiated for specific objectives and which aims atminimizing power consumption through server consolidation, while satisfying a varying load pattern. In this technique a fair allocation CPU resources are givento clients.

Preference-based

In this technique, a market driven auction mechanism is designed to recognize users for resource allocation based on their payment capacities and implements a payment strategy based on a buyer's service preferences. This is a demand-based periodic technique in which cloud service provider auctions cloud resources (VMs) to users and charges them for the amount of resources used.

Comparison of resource allocation techniques*:*

The Various resource allocation techniques are discussed in Table 2.1. Various resource allocation techniques as compared according to metrics discussed above. Which helps the user to decide which the effective Technique for efficient Resources resource allocation and utilization in cloud is computingBased on Reducing Response time and improving performance, scalability and Throughput by proposing rule based technique.

III. Conclusion

Cloud computing is a beneficial for the modern technological world .It is one of most essential technology needed to share and allocate resources online. Cloud is permanent all time storage environments, it is portable and flexible to save and share the resources .In this paper,the analyze is done on various resource allocation techniques used in a cloud environment. It also gives a brief review about resource allocation in the cloud with future enhancements. Based on performance, scalability and availability of the resource.An

efficient Resource Allocation Technique proposed to meet the criteria's likeQuality of Service (QoS) aware utilization of resources, cost reduction and power reduction / energy reduction. The ultimate goal of resource allocation in cloud computing is to maximize the profit for cloud providers and to minimize the cost for cloud consumers

Reference

1. 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT] 978-1-5090-1277-0/16/$31.00©2016 IEEE E. Iniya NehruSenior Technical DirectorNational Informatics Center,Chennai IndiaRanjithBalakrishnanAssistant Professor,TIFAC – CORE in PCT,Velammal Engineering College,Chennai, India
2. 2016 international Conference on Cloud and Autonomic Computing 978-1-5090-3536-6/16 $31.00 © 2016 IEEEDOI 10.1109/ICCAC.2016.16100104 119

Oleg VasilenkoIndustry PracticesEricsson M oscow, Russian Federation

3. Amazon Web Services LLC. Amazon Elastic Compute Cloud (Amazon EC2). http://aws.amazon.com/ec2/, 2009.

4. Amazon Web Services LLC. Amazon Elastic Map Reduce. http://aws.amazon.com/elasticmapreduce/, 2009.

5. AmazonWeb Services LLC. Amazon Simple Storage Service.http://aws.amazon.com/s3/, 2009.

6. D. Battr´e, S. Ewen, F. Hueske, O. Kao, V. Markl, and D. Warneke. Nephele/PACTs: A Programming Model and Execution Framework for Web-Scale Analytical Processing. In SoCC '10: Proceedings of the ACM Symposium on Cloud Computing 2010, pages 119– 130, New York, NY, USA, 2010. ACM.

7. R. Chaiken, B. Jenkins, P.-A. Larson, B. Ramsey, D. Shakib, S. Weaver, and J. Zhou. SCOPE: Easy and Efficient Parallel Processing of Massive Data Sets. Proc. VLDB Endow., 1(2):1265– 1276, 2008.

8Yuxiang Wang, Junzhou Luo, Aibo Song, Fang Dong, "A Samplingbased
Hybrid Approximate Query Processing System in the Cloud",
43rd International Conference on Parallel Processing, IEEE, 2014

9. Avinash Mehta, Mukesh Menaria, Sanket Dangi and ShrishaRao,"Energy Conservation in Cloud Infrastructures",IEEE(2011).

10. Hao Li, Jianhui Liu, Guo Tang," A Pricing Algorithm forCloud Computing Resources", International Conference onNetwork Computing and Information Security, IEEE (2011).

11. Janki Akhani, Sanjay Chuadhary, Gaurav Somani,"Negotiationfor Resource Allocation in IaaS Cloud", ACM,COMPUTE'11,March 25-26, Bangalore, India(2011).\

12. Jianfeng Yan, Wen-Syan Li, "Calibrating Resource Allocationfor Parallel Processing of Analytic Tasks", IEEE InternationalConference on e-Business Engineering, IEEE (2009).

13. Jinhua Hu, Jianhua Gu, Guofei Sun, Tianhai Zhao," AScheduling Strategy on Load Balancing of Virtual MachineResources in Cloud Computing Environment", InternationalSymposium on Parallel Architectures, Algorithms and
Programming,IEEE(2010).

14. Karthik Kumar, Jing Feng, Yamini Nimmagadda, and Yung-Hsiang Lu," Resource Allocation for Real-Time Tasks usingCloud Computing", IEEE (2011)

15. Kazuki MOCHIZUKI and Shin-ichiKURIBAYASHI,"Evaluation of optimal resource allocationmethod for cloud computing Environments with limited electricpower capacity", 2011 International Conference on Network-Based Information Systems, IEEE(2011).

16. Salimath N., Mallappa S., Padhy N., Sheetlani J. (2020) "Scrambling and Descrambling of Document Image for Data Security in Cloud Computing". Innovation, Systems and Technologies, vol 160. Springer, Singapore