International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 11, November-2014

Survey on Workload Management in Cloud

M. Arun kumar
Master Of Engineering
Department of CSE
Angel College of Engineering and Technology
Tirupur, India

S. Selvi
Assistant Professor
Department of CSE
Angel College of Engineering and Technology
Tirupur, India

Dr. B. Kalaavathi
Professor
Department of CSE
K.S.R Institute for Engineering & Technology
Tiruchengode, India

***Abstract*— Cloud computing is mainly used for on-demand services providing. It provides the resource when user's needs. Resources scales up and scales down based on the dynamic needs of uses. Managing the scalability in the cloud is an important issue. Resources are provided to users by two ways. One is static resource provision and the other is dynamic resource provision. Static resource provision can be applied for rare users because day to day life the usage of resources is increased. To manage the scalability only solution is the dynamic resource allocation. Dynamic request taken care by the proactive workload management and the reactive workload management. Proactive workload management is working based on the history of previous years. It is used to satisfy the user's needs based on the history basis of the resource allocation. But one limitation for this proactive workload management is histories are suitable only for few occasions. Rarely histories may not applicable for dynamic workload management. To improve the efficiency of the dynamic workload management some methodologies were proposed. We can survey here how many efficient techniques are used for improving the performance of the workload management**.

Keywords*— Cloud computing, dynamic resource allocation, proactive workload management, dynamic resources

I. INTRODUCTION

Cloud computing is widely used for the scalability in the many services. Here scalability is providing by the dynamic resources management in cloud services. In the static workload management is not efficient for improving the scalability in cloud computing. Cloud computing provides the resources as the services in the cloud application through this we can get the resources from anywhere from the world. Cloud application's scalability is depends upon environment which is used by the users. The cloud application is like providing online shopping means it will popular only in that time of festivals. We cannot assure that same festivals celebrated by all in the world. So scalability predications based on the festivals is an important issue. To improve this predication of scalability many predication algorithm used. This predications used sometimes only. Still we cannot assure the accurate resources measurements. Resources are providing through the virtualization. Virtualization is important concept used to provide the dynamic resources. Load balancing is important techniques to improve the workload management. Load balancing is managing the work load by the dynamic resources. Resources are like CPU, RAM, and Memory. Resources are request by the users. These requests are managed by the service providers. Workload is identified by the threshold. Workload is classified as two types. One is base workload. Second is flash workload. Base workload is defined as the within the threshold values. It is the basic work can be done with these static resources. It does not require the additional resources. Flash work load is defined as crossing the threshold value. In the flash crowd cannot manage the workload. Here the static resources are not enough to manage the flash crowd. Need the additional resources to manage the flash crowd. Dynamic resource management is the one of the solution to manage the resources. To predicate the resources requirements in the dynamic resources allocation data popularity is identified. It is one of the new mechanisms to manage the work load data

II. LOAD PREDICTION AND HOT SPOT DECTECTION

The load prediction algorithm is used to finding the workloads advance in real time. It predicts the workload in given time stamp, but the workload prediction is difficult for making the successful cloud computing. The measurement of serve workloads are the problem of more important in few years, number of prediction techniques are developed for this problem but every workload prediction techniques have some constraints. The load prediction algorithm having two types of approaches:

- Cubic spline Interpolation
- Hot spot detection algorithm

The cubic spline interpolation method will predict the workload. The hot spot detection algorithm is used to measure the virtual machine. The measurement was released to RAM-capacity and CPU and network bandwidth. The load prediction algorithm is depending on previous window size of prediction, function near future workloads. The typical global platform includes the load prediction algorithms. The measurement of raw data is viewed by the implementation of load tracer. The sampling of workload techniques and measurement of workload are taken for load prediction. The workload prediction method for cloud computing is not

efficient for the cloud users. The workload of cloud involves a large file, the caching and characterization of workload will be used to increase the level of performance in the cloud computing. The measurement of future resource was not related to existing measurement. Finally, the load prediction algorithm is reducing the overload problem in the cloud computing environment and it makes the efficient workload management partially.

III. SKEWNESS ALGORITHM

The cloud users are easily increased and decrease the range for resource usage for their needs. The skewness-algorithm measure the server utilization and the algorithm are used to combine the workload and it will make the efficient utilization of resources in the server. The overload avoidance are the main objective of the skewness algorithm and then it will use the green computing. The green computing are satisfies the virtual machine's needs in the cloud environment. The hot spot and cold spot are easily find by the skewness algorithm in cloud environment each and every server having the threshold value for measure the workload for efficient workload management. If the server utilization is higher than the hot-threshold means the server will be denoted of hot-spot if the server resource usage is lower than cold threshold means the server is denoted as cold spot. Each and every resource in the cloud having threshold values. To migrate the particular virtual machine for decrease temperature of server the above hot spot and cold spot method are useful to make the smart workload management in the cloud platforms.

VI. FEATURE LOAD PREDICTION OF VM FUSED

The virtualization method is most important technique in the cloud platform. The cloud system's resources will be multiplexed, that will be used to avoid the overloading problem in each physical machine. The Dynamic allocation of resources will be done by the virtualization technique. The virtualization techniques are used to manage workloads in cloud environment efficiently, The FUSD-algorithm are depending on the existing statistics for workload management and the physical's machines upcoming workloads will be measured by aggregation of resource demands in all virtual machine in physical machine. The CPU-workload will be measured by the FUSD-algorithm in the server. The number of algorithms for load prediction does not provide the actual result, in particular time, it will give peak value when compared to the actual result that will affect the workload management technique, same time, and it will make the overloaded problem that will appear in another time of interval. This algorithm is predicting the overloaded physical machine in the cloud platform. The FUSD Algorithm is the best algorithm for finding virtual machine's workload in the future. When compared to existing methods, The CPU-workload also considered in the algorithm. Finally, the FUSD algorithm are achieving the load balancing techniques and avoid the overloaded problem in the future by the predicting method some time, it will not efficient to the workload management. The Fast Up and Slow Down (FUSD) algorithm are useful for the manage workload in cloud environment efficiently.

V. WORKLOAD TRAILORED ELASTIC COMPUTE UNITS

Dynamic Programming based Scheduling Algorithm is making the successful throughput by to combine the resource of private and public cloud computing. That will make the workload function easily. The WECU are finding the virtual machine instances. The WECU is useful for scientific oriented workloads. This algorithm makes the workload in scalable. The Dynamic type of workload is supported by the WECU. The migration functions of the user from one node to next not increase the computing level. The Dynamic workload is supported by the application level. The workloads characterized by WECU units. The workload is split into two types, one is normal workload and another workload is larger workload that type of workloads is easily handled by the load balancing techniques. The workload are handled by the history based and the application based. The resource computing was done by the workload tailored Elastic compute units. That will make the efficient way to workload management process.

VI. MEASUREMENT , MODELING AND ANALYSIS VIDEO SHARING SITE WORKLOAD

The massive-workload is divided into two types. One is well behaving workload and second one is tres-passing workload. The over-provisioning method was used in the internet based application. The workload scheduling function are making the successful workload management function. The small and base workload are in previous period and used for proactive workload management The well behaving workload are based upon the workload size, If the size is presented in normal and small means, it will be consider into well behavior workload. The trespassing workload was in predictable, that type of workload in sudden spikes. The service level agreements are included into manage the workload in the cloud environment, the interest based Applications are divided based on the workload size. The cloud platforms includes the various types of workloads, the number of workload techniques are proposed for achieving the efficient workload management for the cloud's users and it will increase the performance level and it will make QoS. In good level, the massive workload's technique also provides the way to manage workload in the cloud environment.

VII. ROUND ROBIN ALGORITHM

The round robin algorithm is done with efficient workload. The workload is distributed to all processor by Round Robin order. This round robin algorithm is used for achieving the cost effectiveness. The allocation of workload to the processor are managed by the remote processors, Round Robin algorithm are scalable and flexible for the workload management. The distributions of the workload are same but job processing time is different. The workload is having some approaches that is centralized approach, Distributed approach

and mixed approach. The centralized approach is defined as the single system is to distribute the workload to all systems in the cloud environment. The distributed approach is suitable for the distributed systems. Mixed approach is the combination of the centralized and distributed approaches. These type approaches are used for efficient workload function. The round robin algorithm is more used in the web server application.

VIII. RANDOMIZED ALGORITHM

The Randomized algorithm is currently used in cloud. The randomized algorithm is achieving the scalability and flexibility. The randomized algorithm is a static approach, in this approach; it describes the implementation method and design of system. The randomized algorithm doesn't manage the deterministic approach. The randomized algorithm performs well for equal workload process. The variations of computational are made in the problem in of the system. The remote processor is used to allocate the process for each processor. That will make the efficient workload.

IX. MIN-MIN ALGORITHM

In cloud computing, there are number of algorithms used for workload management. Min-min algorithm is used for un-assignment load. Thus, the minimum value is taken. Load's time is used to schedule the workload to current system machine. The minimum algorithm is making the star vacation problem. The min-min algorithm is to improve the performance level with in the cost. The changes of the system are accepted by the min-min algorithm, because maximum algorithms are more flexible and scalable for the cloud environment.

IX. TOKEN ROUTING ALGORITHM

The token routing algorithm is provides the efficient decision for routing. The distribution of the workload for the agents is not fixed. In Token Routing algorithm, the cloud agents are no need for complete knowledge for the working load. The Token Routing algorithm is used to reduce the system cost. Token Routing algorithm is more scalable when compared to other algorithms. The routing technologies are making the efficient workload. The energy management is not efficient in the Token routing method.

X. CONCLUSION

In this paper we have survey on Workload management techniques. There are many algorithm used to manage the workload in cloud. Load prediction and hot-spot detection is used to predict the workload and measure the VMs. Skewness algorithm used to combine the workload and utilize the resources efficiently. Fast up and slow down algorithm used to improve the efficiency of workload by the concept of VMs techniques. Dynamic Programming based scheduling algorithm is making the successful through put. Massive workload function provides the efficient way to manage the workload. Round robin provides the scalable and flexible for workload management. Randomized algorithm is used for equal workload process. Min-Min algorithm is used to manage the unassigned workload. Token Routing algorithm provides the efficient decision for routing and it is used to efficient workload management.

REFERENCES

[1] Baryshnikov, Y., Coffman, E., Pierre, G., Rubenstein, D., Squillante, M. and Yimwadsana, T. "Predictability of Web-Server Traffic Congestion," 10th International Workshop on Web Content Caching and Distribution (WCW'05), Sophia Antipolis, France.

[2] C.Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A scalable application placement controller for enterprise data centers," in Proc. Of the International World Wide Web Conference (WWW'07), May 2007.

[3] Zhen Xiao, Weijia Song, Qi Chen." Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment".Ieee transactions on parallel and distributed systems, vol. 24, no. 6, june 2013.

[4] X. Kang, H. Zhang, G. Jiang, H. Chen, X. Meng, and K. Yoshihira, "Measurement, modeling, and analysis of Internet video sharing site workload: a case study," in Proc. 2008 IEEE International Conference on Web Services, pp. 278–285.

[5] A. Brian. Load Balancing in the Cloud: Tools, Tips, and Techniques. A Technical white paper in Solutions Architect, Right Scale.

[6] TianShaoliang, Zuo Ming and Wu Shaowei, " An improved load balancing algorithm based on dynamic feedback", Computer Engineering and Design.

[7] hengQi.,et al," Load balancing algorithm based on dynamic feedback", Computer Age, pp: 49-51,2006.

[8] Ren, X., R. Lin and H. Zou, "A dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast" International Conference on. Cloud Computing and Intelligent Systems (CCIS).

[9] A. Khiyaita, M. Zbakh, H. El Bakkali and Dafir El Kettani, "Load Balancing Cloud Computing: State of Art" , 9778-1-4673-1053-6/12/$31.00, 2012 IEEE.

[10]. ZhongXu, Rong Huang,(2009)"Performance Study of Load Balanacing Algorithms in Distributed Web Server Systems", CS213 Parallel and Distributed Processing Project Report..