ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

WORKLOAD PREDICTION IN CLUSTER BASED LOAD BALANCING IN CLOUD ENVIRONMENT

B.BRINDHADEVI
UG student
Department of Computer Science & Engineering
Kongu Engineering College, Erode
brindhapdm@gmail.com
\

S.BHAGYASHREE
UG student
Department of Computer Science & Engineering
Kongu Engineering College, Erode
bhagyas340@gmail.com

A.BHARATH
UG student
Department of Computer Science & Engineering
Kongu Engineering College, Erode
bharathambethkar@gmail.com

Dr.R.S.MOHANA
Assistant Professor(SRG) ,Dept of Computer Science ,
Kongu Engineering College, Erode
mohanaamethesh@gmail.com

KEYWORDS: Workload, Prediction, Clustering, Load Balancing

ABSTRACT

In recent years, the information communication technology (ICT) appeared new paradigm of utility computing called cloud computing. The consumer cloud is always important of high performance for cloud computing service and satisfies Service Level Agreement (SLA). In cloud computing, there is a need of further improvement in task execution time, which will reduce the response time and enhance computing resource utilization. The clustering strategy considers the memory size, million instructions per second (mips) and Ram size of each Virtual machine to realize the optimization for cloud computing environment. The process of host Load is changing over time, so a time series based method is used to predict the future workload based on past workload. The experimental results were conducted in a simulation cloud computing environment. The results show that the virtual machine clustering improves the execution time by reducing the makespan when compared to the execution time of non-clustered virtual machines.

I.INTRODUCTION

Within few years cloud computing grab the IT market very fast and most of the IT industry start using the cloud computing. In cloud computing the word cloud refers as internet, so the meaning of cloud computing is Internet Based Computing. In other words it's a kind of server based computing. Cloud computing provide on demand services to the client. The services includes SaaS (Software as a service) where application software and database access provided to the user pay per use basis, IaaS (infrastructure as a Service) where virtual machine provided to the user using virtualization of physical machine which includes processing power, storage and other resources, PaaS (Platform as a Service) where cloud provider provides a computing platform which includes OS, programming language execution platform and web server.

Growth of cloud computing slower down the efficiency, throughput and utilization of resources for which cloud computing need to be evolved. The prediction model predicts the future workload of cluster of virtual machines. Christo Ananth et al. [5] discussed about Submerge Detection of Sensor Nodes. Underwater networking sensor nodes provide the oceanographic collection of data and monitoring of unmanned or autonomous underwater vehicle to explore sea recourses and gathering of scientific data. The challenges include waterproofing, casing, calibration. Furthermore the research issues are outlined. Based on the predicted workloads load balancing is done on the clusters to improve the execution time The CloudSim toolkit has

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

been used to test the clustering ,prediction and load balancing. The toolkit, a java-based discrete-event cloud computing simulation package, supported both system and behavior modeling of cloud computing system components such as data center and virtual machines (VMs). Mapping of task to resource and resource management are also supported.

II.MATERIALS AND METHODS

Cloud computing service providers have several datacentres in order to optimally serve customer needs around the world. However, existing system does not provide the proper clustering of virtual machines for the cloudlets requested by the customers Every datacenter in cloud computing consist of numerous servers and each server runs numerous VMs. Each cluster has different capability to execute the appropriate cloudlets requested by the customer.

III.CLUSTERING OF VIRTUAL MACHINES

Clustering involves the task of dividing data points into homogeneous clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. In Fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to different clusters. Cluster analysis is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization. It will often necessary to modify preprocessing and parameter until the result achieves the desired properties.

A.FUZZY C-MEANS CLUSTERING (FCM)

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is most widely used. Fuzzy c-means clustering was first reported in the literature for a special case (m=2) by Joe Dunn in 1974. The general case (for any m greater than 1) was developed by Jim Bezdek in his PhD thesis at Cornell University in 1973. It can be improved by Bezdek in 1981. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1.. The FCM algorithm attempts to partition a finite collection of elements $X=\{x_1,x_2,\dots,x_n\}$ into a collection of Fuzzy clusters with respect to some given criterion.

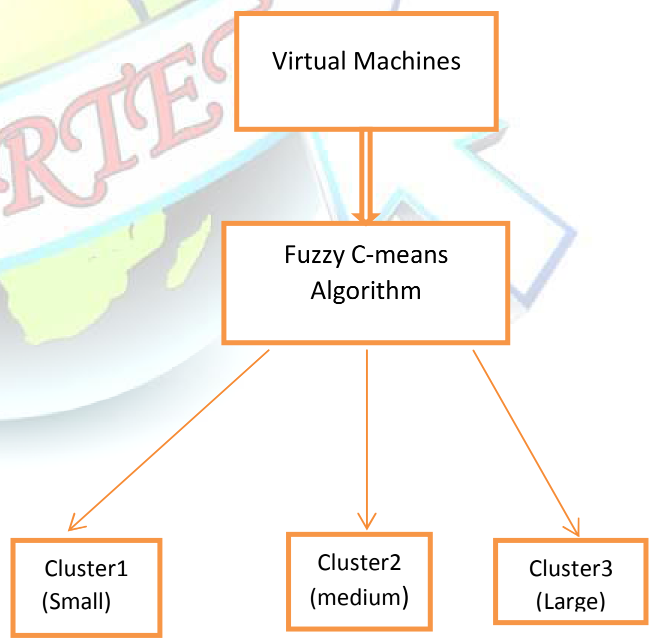


Fig 1:Model for Clustering Of Virtual Machine

In the experiment, the Virtual Machines are clustered into three clusters as small,

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

ALGORITHM 1: FUZZY C-MEAN ALGORITHM

1. Randomly initializing the cluster center
2. Creating distance matrix from a point xi to each of the cluster centers to with taking the Euclidean distance between the point and the cluster center

$$D_{ii}= \sqrt{\sum_{i=1}^{n}(x_i - c_i)}$$

3. Creating membership matrix

$$\mu_{i}(x_{i})=((1/d_{ij})^{1/m-1})/(\sum_{i=1}^{n}(1/d_{ik})^{1/m-1})$$

4. Total membership for a point in sample or decision space must add to 1
5. Generating new centroid for each cluster

$$C_i=(\sum_{i=1}^{n}(\mu_i(x_i))^m.x_i)/(\sum_{i=1}^{n}(\mu_i(x_i))^m)$$

6. Generating new centroid until it matches for two iterations
7. Then finalize the centroid and membership matrix

Here m is real number equal to 1.25, μ_{ij} is the degree of membership of X_i in the cluster ,
x_i is the i^{th} of d-dimensional measured data,
c_i is the d-dimension center of the cluster,

medium and large based on different parameters such as RAM size,Million Instructions Per Second(MIPS) and Size. The cluster is initially assigned with centroid values. The distance matrix is calculated by taking Euclidean distance.Membership matrix is calculated with the help of distance matrix. Total membership for a point in sample or decision space must add to 1. Then the centroid values are calculated. The process is repeated until the centroid values equals the previous iteration's centroid values, and it is taken as final centroid value.Then based on the final centroid value the virtual machines are clustered.The below table gives the algorithm for Fuzzy C-Mean Clustering.

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula.

IV.ARIMA MODEL FOR PREDICTION

ARIMA is AutoRegressive Integrated Moving Average. AutoRegressive is used in statistical calculation in which the future values are estimated based on a weighted sum of past values. Regression analysis is the statistical process for estimating the relationships among variable, it investigates the relationship between a dependent and independent variables. It is used for forecasting and time series modeling . ARIMA model is employed to predict the time series C_k which represents the usage of CPU at time k. In the ARIMA(p,d,q) model, p is the number of autoregressive terms , d is the number of nonseasonal differences and q is the number of lagged forecast errors in prediction equation.

In the experiment, the CPU usage of the cluster of virtual machines for different weeks are retrieved from the past history of

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

usage and the future CPU usage is predicted by using the following steps:

- Moving Average (ma) for the CPU usage(cu) is calculated by using the formula

  ma(i)= cu(i-1)+cu(i)+cu(i+1)/N

  where N= Number of Clusters

- Seasonal Irregular(S_t,I_t) is calculated by

  $S_t I_t(i)$= cu(i)/ma(i)
- Seasonal Component(S_t) is separated from the seasonal irregular
- Regression is performed by using the data analysis tool in which the CPU usage is taken as Y-Axis and time is taken as X-Axis.
- Trend component is calculated by adding the intercept coefficient which was obtained after performing regression and the x variable coefficient and it is multiplied with time t.
- Prediction values are calculated by

  P(i) = S_t(i)/ Trend(i)

V.ARCHITECTURE

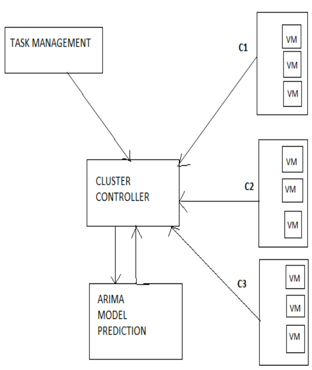


Fig 2: Proposed Architecture

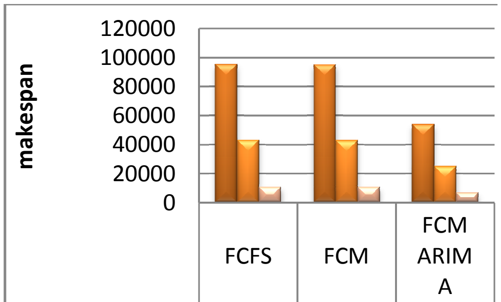
In this architecture, C1,C2,C3 are cluster of VMs. The cluster controller maintains the information about the workload history of clusters. Workload history is given as input to the ARIMA model. The model predicts the future workload and gives it to the cluster controller. The cluster controller then performs the load balancing in the clusters by performing horizontal scaling or vertical scaling based on the predicted workload value.

VI.RESULTS AND DISCUSSION

In order to obtain results of the proposed algorithm the simulation was done using CloudSim 3.0.3 Simulator. In our simulation scenario the proposed algorithm is compared with the existing scenario, for this purpose the following illustrative example is taken. We have created many VMs with different values for the parameters and cloudlets with different length. The cloudlet length varies from 0 to 90000.

In Figure 3, X-axis represents the number of cloudlets and Y-axis represents the makespan of the cloudlets. For Example When number of cloudlets is 30 and virtual machines is 18, the makespan of the FCFS simulation is 10851.300, the makespan after performing FCM is 10837.782,the makespan after load balancing with clustering is 6740.3337. As a result, the makespan is reduced by using clustering with load balancing.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017



	FCFS	FCM	FCM ARIMA
90 cloudlets	95454.27	95266.78	54103.67
60 cloudlets	43107.27	42818.04	25033.60
30 cloudlets	10851.3	10837.78	6740.337

Fig 3: Comparison Chart OfMakeSpan

VII.CONCLUSION

Cloud computing is a distributed computing which mainly focuses on providing services to the customers and it provides computational as well as storage resources to users. To improve the execution time, load balancing with clustering provides the solution. Clustering is the process in which the similar virtual machines are clustered load balancing is the process in which the VMs are either removed or added to the cluster. The proposed architecture aims at achiving the minimum makespan and maximum utilization of the cloud resources.

VIII.REFERENCES

[1]M. Armbrust,"A view of cloud computing", Communications of the ACM, vol. 53, no. 4,(2010), pp.50-58.

[2] Mell, Peter, and Timothy Grance. "The NIST definition of cloud computing (draft)."NIST special publication, vol. 800, no. 145, (2011), p.7.

[3] M.S.Yang," A Survey of fuzzy clustering" Mathl. Comput.Modelling Vol. 18, No. 11, pp. 1 -16, 1993.

[4] Uri Kroszynski and Jianjun Zhou, Fuzzy Clustering Principles, Methods and Examples, IKS, December 1998.

[5] Christo Ananth, S.Surya, Berlin Mary, "Submerge Detection of Sensor Nodes", International Journal Of Advanced Research Trends In Engineering And Technology (IJARTET), Volume II, Special Issue XXV, April2015

[6] G. E. Box, G. M. Jenkins, and G. C. Reinsel, Time series analysis: forecasting and control. *Wiley*, 2011, vol. 734.

[7]P.A. Dinda, D.R. HallaronHost load prediction using linear modelsJ. Clust. Comput., 25 (3) (2010), pp. 265–280.

[8] Hu, R., Jiang, J., Liu, G., Wang, L., 2013. CPU load prediction using support vector regression and Kalman smoother for cloud. In: Proceedings of 33rd IEEE International Conference on Distributed Computing Systems, Philadelphia, USA, pp. 88–92.

[9] Zhou, W., Yang, S., Fang, J., Niu, X., 2010. Vmctune: a load balancing scheme for virtual machine cluster using dynamic resource allocation. In: Proceedings of 9th IEEE International Conference on Grid and Cooperative Computing, Munich, Germany, pp. 81–86.