IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10546

AN EFFICIENT RESOURCE ALLOCATION IN CLOUD COMPUTING USING KNN AND NAIVE BAYES ALGORITHMS

R.M.Sreesantthiya[1], K.Selva Priya[2], Dr.S.Kayalvili M.E.,M.B.A.,Ph.D.,[3]

Final B.E (CSE), Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamilnadu, India[1,2]

Associate Professor(CSE), Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamilnadu, India[3]

Abstract: The amount of computing resources required by current and future data-intensive applications is increasing dramatically, creating high demands for cloud resources. We provide a method in cloud to analyze the resource requirements for the tasks and allocate those resources at right time. In addition, it ensures a stable cloud structure and rapid resource allocation. In the proposed system, the tasks are classified based on its size. The resources are allocated for parallel completion of those by VMs. It uses KNN and Naive Bayes algorithm for task clustering and resource allocation respectively. It is simple to implement, improves flexibility and reduces resource allocation time.

Keywords: Cloud computing, resource demands, KNN(K Nearest Neighbor) and NB(Naive Bayes) algorithms, prediction accuracy, physical machine(PM), virtual machine(VM).

I. INTRODUCTION

Cloud is used for the purpose of storing the data. Generally, cloud computing means an on-demand availability of plenty of computer resources in the system, especially its storage of data (cloud storage) and computing power without active management by a cloud user. This concept is specifically used to describe the data centers available to plenty of users over the Internet. Nowadays, large clouds frequently have functions distributed over several locations from central servers. Sometimes clouds are be limited to one organization (enterprise clouds), or be available to multiple organizations (public cloud). Cloud computing relies on sharing of resources to understand coherence and economies of scale. Public advocates and hybrid clouds point that cloud computing allows many companies to avoid or minimize up-front IT infrastructure costs. Proponents claim that cloud computing allows enterprises to urge their applications up to run faster, with improved ability in management and fewer maintenance, which it enables IT teams to adjust resources to meet unpredictable demand, providing the rapid computing capability: high computing power at certain periods of peak demand. Generally cloud platform provides various types of resources to automatically create VMs This type of platform can flexibly increase or decrease the number of VMs according to changes in resource demands of the user and resource loads. However, a cloud platform carries numerous applications and services, which requires a diverse selection of VMs. In addition to the increased volumes of resource demands can lead to many changes and strong fluctuations and errors in the server load. If the server load changes at a faster rate than that of the migration of VMs on the server, the performance of applications running on these VMs can be affected, or some unnecessary migrations of VMs may cause the cloud platform to become unstable. These type of problems can be rectified by the prediction of future resource demands and server loads.

II. SYSTEM DESIGN

The algorithm enhances the VM selection phase based on real time monitoring data collections and analysis of physical and virtual resources. Our aim is to strengthen VM scheduling .In order to incorporate criteria related to the actual VM utilization levels, so VMs can be placed by minimizing the penalization of overall performance levels. The optimization scheme involves analytics to an already deployed VMs to incorporate maximization of utilization levels and minimization of performance drops. A monitoring engine that permits online resource usage monitoring data collection from VMs. This engine is capable of collecting and collecting data based on regular intervals and stores it in online cloud service that makes it readily available for data processing. Data is collected every time interval (e.g. 1 second)

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10546

and is stored in a temporary local file. Supervised machine learning algorithms are collected for classification purpose and the input dataset is desired to be a labeled one.

K Nearest Neighbor's Method

K Nearest Neighbor's may be a simple algorithm that handles all accessible cases and groups new cases supported a specific similarity measure (e.g., distance functions). KNN also has been used in statistical determination and pattern recognition. The K Nearest Neighbor's algorithm (KNN) is also a non-parametric method generally used for classification and regression techniques.

Euclidean distance= $\sqrt{\sum_{i=1}^{k}((x_1 - x_2)^2 + ((y_1 - y_2)^2 + ((z_1 - z_2)^2)}$

Naive Bayes (NB) method

The Naive Bayes Classifier technique is predicated on Bayesian theorem and is especially used when the dimensionality of the inputs is high. The Bayesian Classifier is capable of calculating the foremost possible output supported the input. It is also possible to feature new data at runtime and have a far better probabilistic classifier. A naive Bayes classifier considers that the presence (or absence) of a specific feature (attribute) of a category is unrelated to the presence (or absence) of the other feature when the class variable is given. For example, a fruit could also be considered to be an apple if it's red, round. Even if these features depend on each other or upon the existence of other features of a category , a naive Bayes classifier considers all of these properties to individualistically contribute to the possibility that this fruit is an apple. Algorithm works as follows,

$$P(\text{label} \mid \text{features}) = \frac{P(\text{label}) * P(\text{features} \mid \text{label})}{P(\text{features})}$$

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

$$P(C|X)=P(X_1|C)*P(X_2|C)* \ldots\ldots P(X_n|C)*P(C)$$

P(c|x) is the posterior probability of class given predictor of class.
P(c) is considered as an initial probability of class.
P(x|c) is the probability of a predictor of the given class.
P(x) is an initial probability of predictor of class.
Bayes theorem paves the way for calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c) and Naive Bayes classifier examines that the effect of a value of the predictor (x) on the given class (c) is independent of the values of the other predictors.

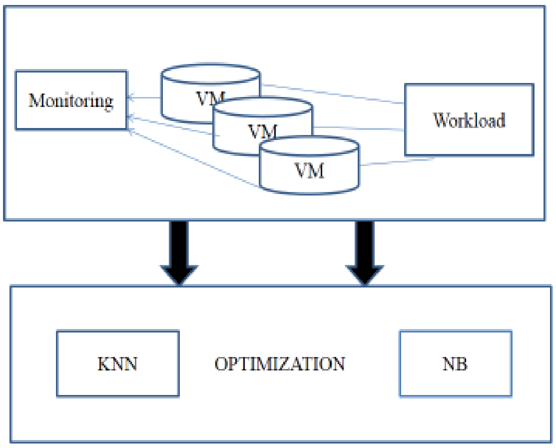

Fig.1 Architectural diagram

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10546

Optimization scheme

Optimization scheme is to define the weight of the PM according to the resource usage of the VMs. This will gives away the information about previously deployed VMs status, like indications that the task is processing or not. To achieve this we provide two optimization schemes. Here classification of the VM status about its current resource usage is classified using the knn and nb. Initially the virtual machine resource usage dataset is collected and monitored and then the collected data is classified using the machine learning methods like KNN and NB.

III. EXPERIMENTAL PRECEDURE

The multi-domain dataset includes the various resource utilization from of cloud resources such as bandwidth, memory, cpu. The entire domain consist of 1000 labeled instances which are assumed here as the past resources utilization history record. The task performed by 28 physical machine when using SVM is reduced by 24 physical machine by using knn &nb classifier algorithm also the error rates gets reduced by 0.025%.

No. of physical machines: 24
No. of virtual machines : 16
No.of classifiers : 02

The absolute error is defined as the absolute value of the difference between the measured value and the true value. Thus, let:

e_a = the absolute error
x_m = the measured value
x_t = the true value

The formula for computing absolute error is:

$e_a = | x_m - x_t |$

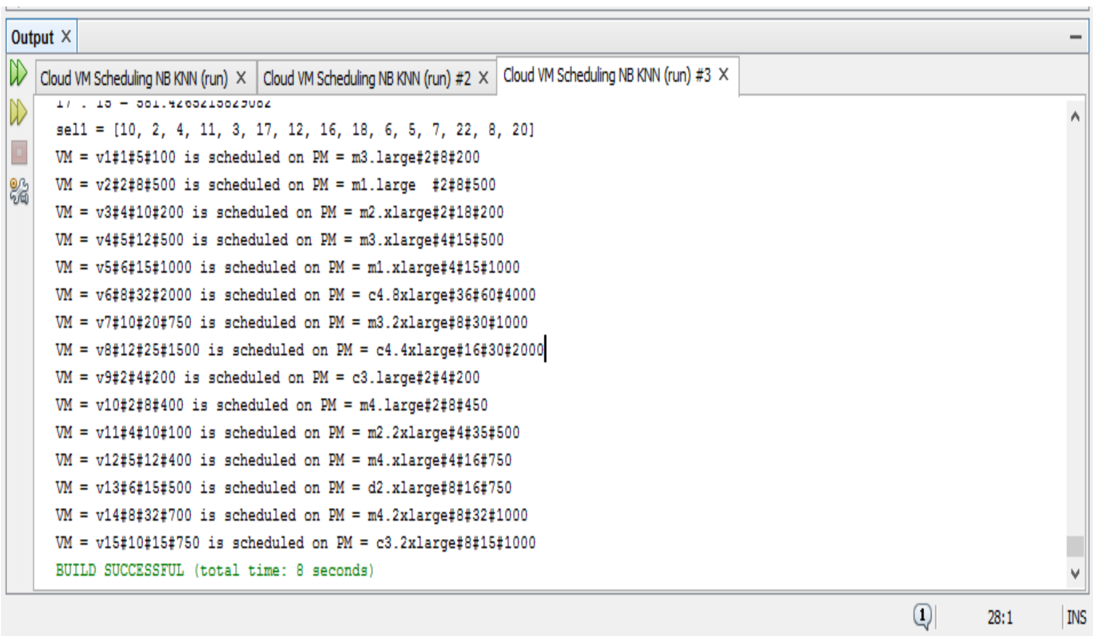

Fig.2 VM scheduling on PM

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10546

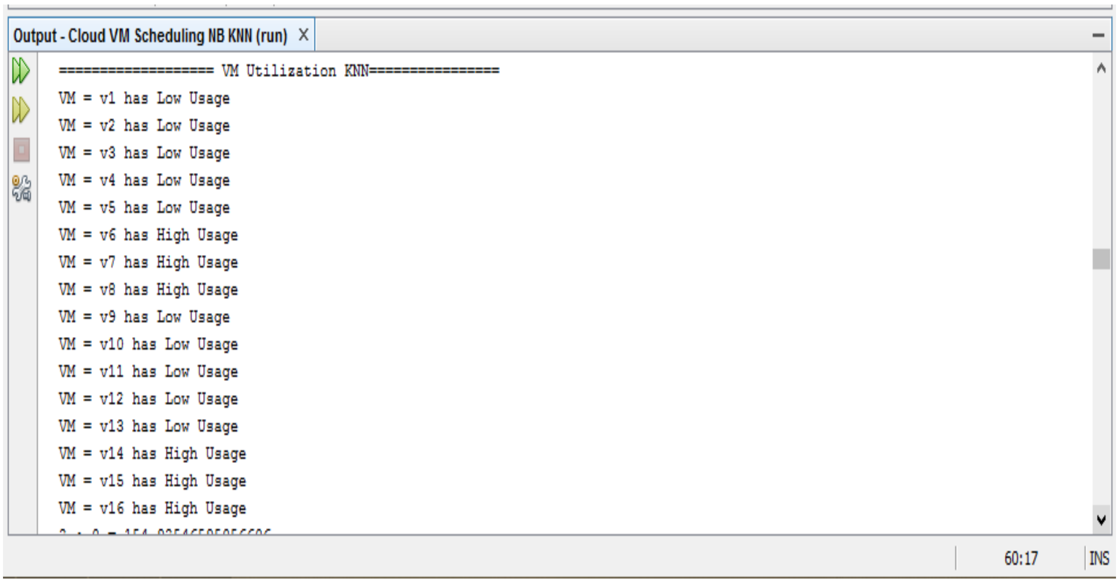

Fig.3 VM utilization using KNN

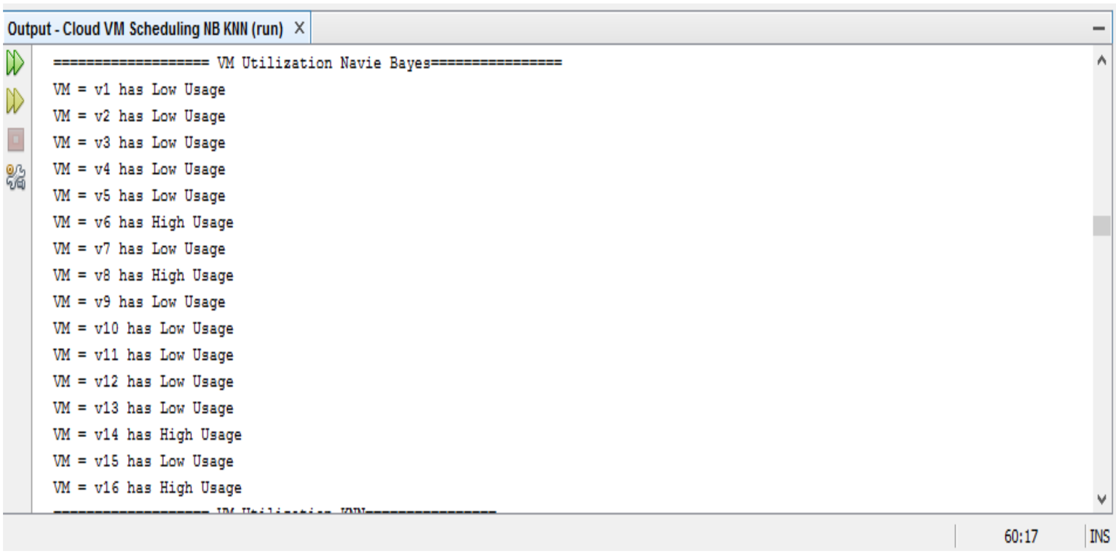

Fig.4 VM utilization using Navie Bayes

The results evaluate the past resource utilization levels and classifies according to the overall resource usage. At the top the list of candidate hosts is populated and therefore the resources are ranked accordingly. In detail, by using this algorithm PMs are re-ranked consistent with the chosen optimization scheme and supported their VM usage.

IV. CONCLUSION

Different virtual machine placement algorithms were used for scheduling by choosing physical machines according to the system data (i.e. usage of CPU, memory, bandwidth) in cloud system. The present VM placement doesn't take into account of real time VM resource utilization levels. Here we a new VM placement algorithm based on past VM usage experiences is proposed then the VM usage is monitored and the data gets trained using machine learning models (KNN&NB) to calculate the prediction of the VM resource usage, to place VMs accordingly. An algorithm that allows VM placement according to PM and VM usage levels and computational learning method based on the concept of analyzing past VM resource usage according to historical records to optimize the PM selection phase was introduced. Also, a VM placement algorithm based on real time virtual resource monitoring was introduced where machine learning models is used to train and learn from previous virtual machine resources usage. Thus, a monitoring engine is assumed

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10546

with resource usage data. The count of the physical machine gets reduced by 4 by using knn & nb classifier than Support Vector Machine (SVM) classifier.

REFERENCES

[1] Jing Chen and Yinglong wang, "An Adaptive Short-Term Prediction Algorithm for Resource Demands in Cloud Computing", IEEE Access. Vol. 8,pp. 53915 – 53930, 2020.
[2] Q. Wu, M. Zhou, Q. Zhu, Y. Xia, and J. Wen, "MOELS: Multiobjective evolutionary list scheduling for cloud workflows," IEEE Trans. Autom. Sci. Eng., vol. 17, no. 1, pp. 166–176, 2020.
[3] J. Chen and Y. Wang, "A hybrid method for short-term host utilization prediction in cloud computing," J. Electr. Comput. Eng., vol.8 , Art. no. 2782349, 2019.
[4] S. Sharifian and M. Barati, "An ensemble multiscale wavelet-GARCH hybrid SVR algorithm for mobile cloud computing workload prediction," Int. J. Mach. Learn. Cybern., vol. 10, no. 11, pp. 3285–3300, 2019.
[5] Q. Wu, F. Ishikawa, Q. Zhu, and Y. Xia, "Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters," IEEE Trans. Services Comput., vol. 12, no. 4, pp. 550–563, 2019.
[6] E. Furuncu and I. Sogukpinar, "Scalable risk assessment method for cloud computing using game theory (CCRAM)", Computer Standards & Interfaces, vol. 38, pp. 44–50, 2015.
[7] J. Li, J.W. Li, and X. F. Chen, "Identity-based encryption with outsourced revocation in cloud computing", IEEE Transactions on Computers, vol. 64, no. 2, pp. 425–437, 2015.
[8] X. Yi, F. Y. Rao, and E. Bertino, "Privacy-preserving association rule mining in cloud computing", in Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, pp. 439–450, 2015.
[9] J. Lou and Y. Vorobeychik, "Equilibrium analysis of multi-defender security games", Proceedings of the Twenty-Fourth International Joint Conference on Artifical Intelligence (IJCAI), pp. 596–602, 2015.
[10] Z. Ismail, C. Kiennert, J. Leneutre, and L. Chen, "Auditing a cloud provider's compliance with data backup requirements: A game theoretical analysis", IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1685–1699, 2016.
[11] N.K. Sharma and A. Joshi, "Representing attribute based access control policies in owl", IEEE Tenth International Conference on Semantic Computing(ICSC), pp. 333–336, 2016.
[12] P. Narwal, D. Kumar, and M. Sharma, "A review of game-theoretic approaches for secure virtual machine resource allocation in cloud", Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, 2016.
[13] Y. Yong, M. H. Au, and G. Ateniese, "Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage", IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, pp. 767–778, 2017.
[14] M. Nabeel and E. Bertino, "Privacy preserving delegated access control in public clouds", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 9, pp.2268–2280, 2014.
[15] U. Wazir, F. G. Khan, S. Shah, "Service level agreement in cloud computing: A survey", International Journal of Computer Science and Information Security, vol. 14, no. 6, p. 324, 2016.