

SMART RESOURCE USAGE PREDICTION USING CLOUD COMPUTING FOR MASSIVE DATA PROCESSING SYSTEMS

Dr. Abraham Chandy,
Professor,
Karunya University,
Coimbatore, India.
Email id: abrahamchandy@karunya.edu

Abstract: Resource management plays the vital role in the cloud computing as the requirement for the massive data processing system such as health sectors, business solutions and the internet of things keeps on increasing in an exponential range. Allocation of proper and perfect resources remains as the main reasons for the successful computation of the applications. However the conventional resources management methodologies, that totally depends on the simple heuristic based methods fails to accomplish a performance that is predictable. The appropriate resource allocation is directly related to the workload demand prediction as the would help to bring down the cost, time and power and the memory usage. The proposed method in the paper leverages the machine learning approaches to manage the resource allocation in the cloud computing for the massive data processing system, the simulation of the proposed model using the network simulator -2 enables to achieve a better performance and resources utilization at a decreased cost, time, power and memory usage.

Keywords: Cloud Computing, Resource Allocation, Work Load Prediction, Massive Data Processing System, Machine Learning Algorithms.

1. INTRODUCTION

The information technology has attained a quick progress, due to the conversion of the traditional mainframe architectures to client server model. Further the access of multiple services via internet has rapidly increased since the emergence of the concept of cloud computing that offers a distributed computing platform in a highly scalable manner with the options of pay as you use and service on demand. the advancement in the network technology, and the bandwidth provisions affording high speed , the seamless communication and internet connection is made possible in any part of the world, initially the cloud computing was used as consumer model , rendering services as per the demands (software, platform or infrastructure) put forth by the user. The enhanced scalability and the flexibility and

the distributed nature of the cloud computing has attracted more and more users and large scale industries to adopt cloud computing, in order to achieve their business goals at a reduced time and cost, this caused a massive set of data flow that requires a proper resource management. The figure.1 below shows the architecture of cloud reference by the NIST.

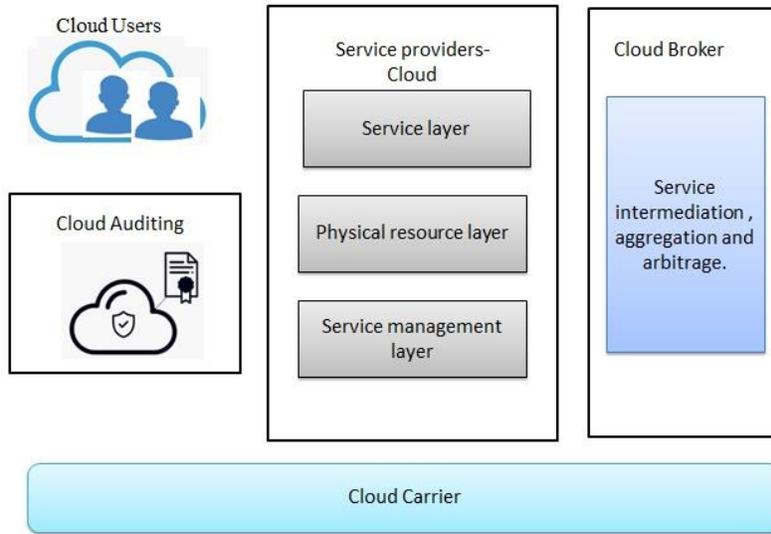


Figure.1 Cloud Architecture

The perfect planning for the resources seems to be an challenging and tedious task in cloud computing due to the continuous changes endured in the demand for the cloud resources and the resource allocation utilizing conventional methodologies are failure as the performance of the system are in predictable. The cloud computing is modeled to offer computing, storage and network resources required as service on demand, with its different mode of deployments, such as private, public , hybrid and community cloud. The general resource management process includes discovery of resources, allocation and observing of the process. The process manages the physical resources such as the processing power, memory, cost and time of execution etc. The figure.2 below shows the conventional resource allocation process.

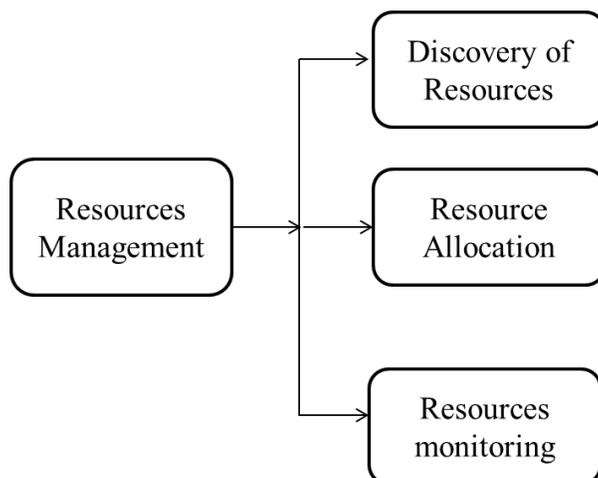


Figure .2 Resource Allocation Process

Most of the resource management process just allocates the results based on the demand without predicting the load of the demand, so the resources allocated are at certain times incapable of completing the work and at certain times work load allotted is very low but charged for the entire duration, increasing the resource utilization cost. So this makes it necessary to predict the work load before the resources are allotted to it.

The method in the paper follows a resource allocation method based on the work load prediction so as to bring down the cost, time, memory and the power utilization for the massive data processing system. The proposed method employs the machine learning approaches to predict the work load as well as allocate the resources. The rest of the paper is organized with the section 2 presenting the related works of the existing methods, section 3 presenting the proposed resource allocation, section4 presenting the performance analysis and section 5 presenting the conclusion followed by references.

2. RELATED WORKS

Work flow scheduling and resource allocation being very tedious tasks in the cloud computing due to the changing demands and the requirement to render service to the users in an acceptable manner, the smart resource management

is essential. The related works in the section presents the measure taken so far to manage the resources for a massive data processing system. Author Kumar, D. et al [1] in his paper presents the “review on the task scheduling in the clouds environment that is ubiquitous” Bashar, Abul et al [2] presents the cloud computing based management to handle the massive data flow of the industry.

The author Bhalaji, N. et al [3] has proposed the task scheduling and the resource allocation that is delay diminished for better resource management heterogeneous cloud environment, further the utilization of the cloud in managing the huge set of data flow from the smart connected objects was proposed by the author Duraipandian, M., et al [4] mean while the author Mohamaddiah, Mohd et al [5] presents the survey on the resource allocation as well as monitoring in cloud paradigm and Bashar, Abul et al [6] proposes the offloading techniques to manage the mobile cloud computing securely in accost efficient manner. Smys, S et al [7] proposes the "Self-organizing hierarchical structure for wireless networks."

Younge, Andrew et al [8] proposes the efficient techniques to manage the resources in the cloud computing paradigm and ensuring sustainability by enabling a green computing and the Suma, V. et al [9] has also proposed a big data analytics frame work clubbing the cloud computing to achieve sustainability in industrialization. Demirci, et al [10] presents a survey on the machine learning techniques to provide an energy efficient resource management in the cloud computing platform.

Jennifer S. Raj et al [11] has proposed the frame of efficient work load management in cloud integrating the big data analytics and the internet of things for a health care sector. Ramesh, S., et al [12] proposes the "Performance analysis of heuristic clustered (HC) architecture in wireless networks." Kumar et al [13] has proposed the Workload prediction in cloud using artificial neural network and adaptive differential evolution." Hu, et al [14] proposes the "Workload prediction for cloud computing elasticity mechanism." and the author Sahi, Supreet Kaur, et al [15] elaborates the review of the work load prediction for the cloud services.

3. PROPOSED WORK

As the resource management totally depends on the job demands, it was identified that the work load prediction would improve the resource allocation process in terms of time, cost, memory and power utilization , further to automate the process of prediction and the resource allocation the proposed method leverages the machine learning approaches that learns the behavior of the previous jobs to predict the workloads , the workloads predicted are further prioritized based on the deadline of the jobs and the proposed method employs weighted quadratic random forest algorithm that is based on the traditional random forest to predict the work load and utilizes the genetic algorithm to allocate the appropriate resources to the jobs. The figure.3 below shows the over view of the proposed resource management method.

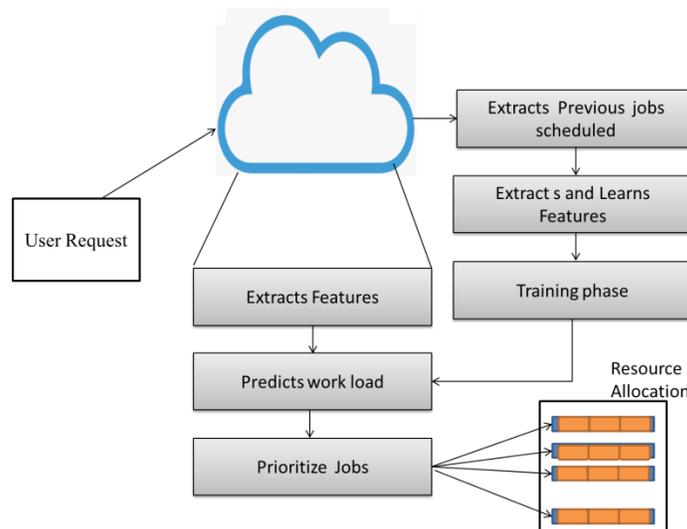


Figure.3 Proposed Resource Management

The method followed utilizes the weighted quadratic random forest algorithm to predict the work load and prioritize the jobs. Once the user request is send from the user to the cloud the features of the work in terms of the memory

requirement, the deadline for the job, the time of arrival, and the computation power required to perform the process are gathered, the table.1 below show the acronyms used.

Parameters	Acronyms
Time of arrival	A_T
Expiry time	E_T
Memory Requirement	M_R
CPU Utilization	$CPU_{utilization}$
Network Utilization	$Net_{utilization}$
Waiting requests	$wait_{req}$
Terminated/blocked	req_{bloc}
Running requests	run_{req}

Table.1 Acronyms used

Once the features are extracted work load is predicted applying the training data observed from the learning phase that learns the features of the job from the previous workloads scheduled, the figure.4 below provides the detail of the features extracted from the user request.

User identification $\{U_1, \dots, U_n\}$	Request number $\{R_1, \dots, R_n\}$	Time of arrival $\{A_T\}$	Expiry time $\{E_T\}$	Memory Requirement $\{M_R\}$
--	---	------------------------------	--------------------------	---------------------------------

Figure.4 Extracted Features
113

Based on the deadline the jobs are prioritized and sent for resource allocation. The algorithm below in figure.5 shows the prioritization of jobs.

```
Input : user request  $\{U_1, \dots, U_n\}$ 
Output: Prioritized Request  $\{R_1, \dots, R_n\}$ 
Start
Accept user requests
Initialize state  $s=0$ 
For  $s=0$ 
Extract features  $A_T, E_T, M_R$ 
Examine features
Enumerate  $E_T$ 
If  $(E_T(U_n) < E_T(U_{n-1}))$ 
Replace the position of  $E_T(U_{n-1})$  with the
position of  $E_T(U_n)$ 
Update position
Increment S
If  $(S=S_n)$ 
Stop
else
Next S
End
```

Figure.5 Prioritization of Jobs

The random forest algorithm used extracts the features employing the bootstrap resampling technique and develops a decision tree for every feature observed and classifies the decision tree implementing simple vote and with the large vote of classification and provides the final result of prediction, the training works in parallel, to significantly enhance the efficiency, k number of decision trees are combined and trained in the same way and the input user requests are now classified depending on the simple voting of the outcomes of the every decision tree. Further the genetic algorithm (GA) is applied to select the appropriate resources by subjecting the resources through the process of selection, crossover and mutation. The resources satisfying the needs of the user request are identified and allotted; this process is done for every user request that is sent to the cloud.

4. RESULTS ANANLYSIS

The proposed method is evaluated using the workflow simulator to evince the enhancement in the performance in terms of the following parameters, CPU , memory, disk and network utilization and the system level features that include the waiting requests, running requests, request terminated/blocked due to the insufficient resources. Figure .7 (a) and 7. (b) Shows the percentage of the CPU, memory, disk and network utilization and the system level features and its comparison with the conventional methods.

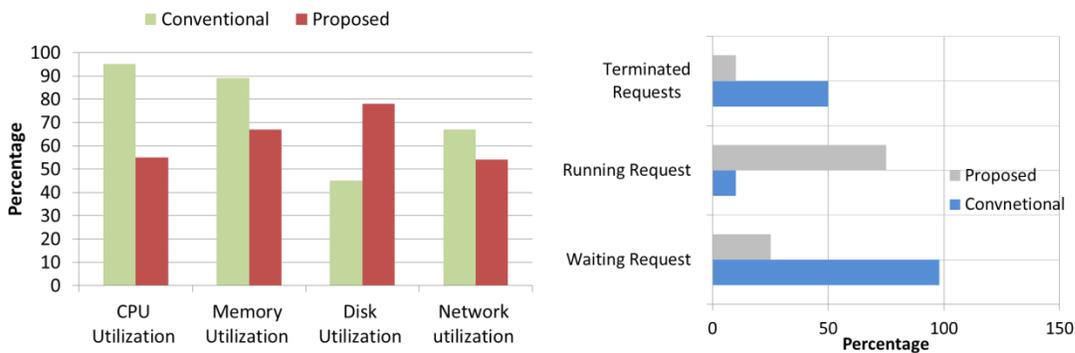


Figure.7 (a) (left) Hardware utilization and (b) (right) System Level features

Where the

$$CPU_{utilization} = CPU_{idle} + CPU_{used} + CPU_{speed} \quad (1)$$

$$Net_{utilization} = sent_{bytes} + Received_{bytes} \quad (2)$$

Where disk utilization gives the particulars of the amount of free space and the memory utilization shows the amount of memory used.

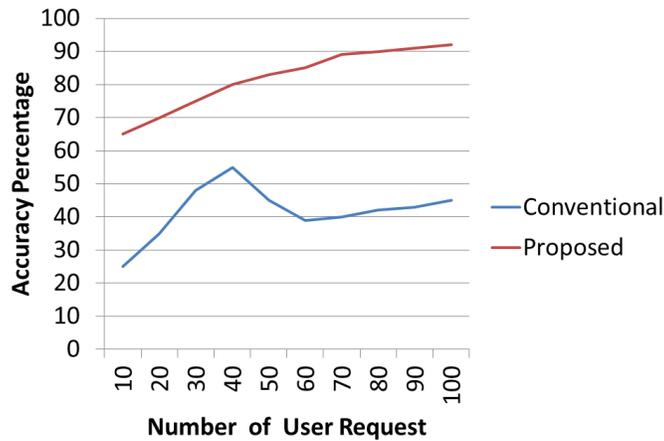


Figure.8 Accuracy in prediction and Allocation

The figure. 8 above shows the accuracy in work load prediction and the resource allocation for varying number of requests, and its comparison with the conventional heuristics methods. Further the table below provides the CPU, memory, disk and network utilization and the system level features for varying number of user request.

No. of User Request	CPU utilization %			Memory Utilization %		Network Utilization %		Disk Utilization%	
	Idle	Used	Speed	used	Un used	No. of send	No. of Received	Used	Un used
10	45	55	89	45	55	60	56	30	70
20	34	66	88	44	56	67	65	35	65
30	32	68	87	34	66	78	75	37	63
40	30	70	90	36	64	74	73	45	55
50	25	75	92	40	60	65	63	55	45
100	35	65	94	50	50	80	72	59	41

Table.2 Performance Analysis

5. CONCLUSION

The smart resource management in the cloud for the massive data processing system is accomplished in the paper leveraging the machine learning. As the work load demand is connected with the resource allocation, the proposed method predicts the work load using the random forest and assigns the resources using the genetic algorithm. The results observed evinces the capability of the proposed method in terms of the resource utilized, system level features and the accuracy observe din prediction. The paper in its future is focused to concentrate on the security issue that has to be addressed during the allocation of the results.

References

- [1] Kumar, D. "Review on task scheduling in ubiquitous clouds." *J. ISMAC* 1, no. 01 (2019): 72-80.
- [2] Bashar, Abul. "INTELLIGENT DEVELOPMENT OF BIG DATA ANALYTICS FOR MANUFACTURING INDUSTRY IN CLOUD COMPUTING." *Journal: Journal of Ubiquitous Computing and Communication Technologies September 2019*, no. 01 (2019): 13-22.
- [3] Bhalaji, N. "DELAY DIMINISHED EFFICIENT TASK SCHEDULING AND ALLOCATION FOR HETEROGENEOUS CLOUD ENVIRONMENT." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 01 (2019): 51-62.
- [4] Duraipandian, M., and Mr R. Vinothkanna. "Cloud based Internet of Things for smart connected objects." *Journal of ISMAC* 1, no. 02 (2019): 111-119.
- [5] Mohamaddiah, Mohd Hairy, Azizol Abdullah, Shamala Subramaniam, and Masnida Hussin. "A survey on resource allocation and monitoring in cloud computing." *International Journal of Machine Learning and Computing* 4, no. 1 (2014): 31.
- [6] Bashar, Abul. "SECURE AND COST EFFICIENT IMPLEMENTATION OF THE MOBILE COMPUTING USING OFFLOADING TECHNIQUE." *Journal of Information Technology* 1, no. 01 (2019): 48-57.
- [7] Smys, S., G. Josemin Bala, and Jennifer S. Raj. "Self-organizing hierarchical structure for wireless networks." In *2010 international conference on advances in computer engineering*, pp. 268-270. IEEE, 2010.

- [8] Younge, Andrew J., Gregor Von Laszewski, Lizhe Wang, Sonia Lopez-Alarcon, and Warren Carithers. "Efficient resource management for cloud computing environments." In *International Conference on Green Computing*, pp. 357-364. IEEE, 2010.
- [9] Suma, V. "TOWARDS SUSTAINABLE INDUSTRIALIZATION USING BIG DATA AND INTERNET OF THINGS." *Journal of ISMAC* 1, no. 01 (2019): 24-37.
- [10] Demirci, Mehmet. "A survey of machine learning applications for energy-efficient resource management in cloud computing environments." In *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*, pp. 1185-1190. IEEE, 2015.
- [11] Smys, S., and Jennifer S. Raj. "INTERNET OF THINGS AND BIG DATA ANALYTICS FOR HEALTH CARE WITH CLOUD COMPUTING." *Journal of Information Technology* 1, no. 01 (2019): 9-18.
- [12] Ramesh, S., and S. Smys. "Performance analysis of heuristic clustered (HC) architecture in wireless networks." In *2017 International Conference on Inventive Systems and Control (ICISC)*, pp. 1-4. IEEE, 2017.
- [13] Kumar, Jitendra, and Ashutosh Kumar Singh. "Workload prediction in cloud using artificial neural network and adaptive differential evolution." *Future Generation Computer Systems* 81 (2018): 41-52.
- [14] Hu, Yazhou, Bo Deng, Fuyang Peng, and Dongxia Wang. "Workload prediction for cloud computing elasticity mechanism." In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pp. 244-249. IEEE, 2016.
- [15] Sahi, Supreet Kaur, and V. Dhaka. "A review on workload prediction of cloud services." *International Journal of Computer Applications* 109, no. 9 (2015): 1-4.