a
Journal of Engineering Sciences
Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

FRAMEWORK FOR TASK SCHEDULING IN CLOUD USING MACHINE LEARNING TECHNIQUES

Dr.S.Sree Hari Raju ,Associate Professor,B.tech(CSE),M.tech(PC),Ph.d(CSE),sreehariraju.s@nnrg.edu.in

M.Jagdeesh kumar, B.tech, Department of CSE, jagdeeshmakula007@gmail.com

P.Nikhil, B.tech, Department of CSE, nikhilpalleboina129@gmail.com

G.Sandeepraj, B.tech. Department of CSE, guggillasandeepraj@gmail.com

ABSTRACT: Task scheduling plays a vital role in the function and performance of the cloud computing system. While there exist many approaches for improving task scheduling in the cloud, it is still an open issue. In this proposed framework we try to optimize the utilization of cloud computing resources by using machine learning techniques. Task scheduling algorithms can be designed for static or dynamic scenarios. The proposed framework is for the dynamic scenario. Task scheduling can consider different parameters for scheduling purposes like Makespan, QoS, energy consumption, execution time, and load balancing. We propose to apply a machine learning technique for the incoming task requests so as to classify the best suitable algorithm for the task request rather than randomly assigning the scheduling algorithm. Supervised machine learning techniques can be used here. The outcome of the proposed work leads to the selection of the best task scheduling algorithm for the input task(request).

Keywords: *railway track crack detection, deep learning models*

1. INTRODUCTION

Within no time machine learning and cloud computing have gained the concentration of the IT industry. With the high availability of the internet at a lower cost and an upcoming enormous number of the app's data generated is also a vast amount. Cloud computing delivers data and resources on a pay-per-usage basis. Scheduling processes, tasks, and resources like CPU, memory, and peripherals is the major concern in cloud computing. Task scheduling in a cloud environment can be done statically or dynamically. The parameters to be considered for task scheduling in a cloud environment can be listed as execution time, energy consumption, response time, cost, makespan, QoS, etc. machine learning is another area of the IT industry which is playing a very vital role in providing better services and solutions to IT customers by enabling smart services. The key idea here is to combine machine learning techniques with task scheduling in a cloud environment. The machine learning always provides smart services to the end-users. The machine learning technique either supervised or unsupervised method can be used externally for the decided parameters. It learns and classifies which algorithm is best for the

scenario, based on the classification the request passed on to the corresponding data center.

2. LITERATURE REVIEW

2.1 Cloud computing: stateof-the-art and research challenges

Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Internet. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand. However, despite the fact that cloud computing offers huge opportunities to the IT industry, the development of cloud computing technology is currently at its infancy, with many issues still to be addressed. In this paper, we present a survey of cloud computing, highlighting its key concepts, architectural principles, state-of-the-art implementation as well as research challenges. The aim of this paper is to provide a better understanding of the design challenges of cloud computing and identify important research directions in this increasingly important area.

2.2 Survey on scheduling issues in cloud computing

Cloud computing has captured the attention of today's CIOs, offering huge potential for more flexible, readily-scalable and cost-effective IT operations. It represents a different way to architect and remotely manage computing resources. Cloud computing deals with different kinds of virtualized resources, hence scheduling places an important role in cloud computing. In cloud, user may use hundreds of thousands virtualized resources for each user task. Hence manual scheduling is not a feasible solution. Focusing scheduling to a cloud environment enables the use of various cloud services to help framework implementation. Thus the comprehensive way of different type of scheduling algorithms in cloud computing environment surveyed which includes the workflow scheduling as well as grid scheduling. This study gives an elaborate idea about grid, cloud, workflow scheduling.

2.3 Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment

Cloud computing is an emerging technology in distributed computing which facilitates pay per model as per user demand and requirement. Cloud consist of a collection of virtual machine which includes both computational and storage facility. The primary aim of cloud computing is to provide efficient access to remote and geographically distributed resources. Cloud is developing day by day and faces many challenges, one of them is scheduling. Scheduling refers to a set of policies to control the order of work to be performed by a computer system. A good scheduler adapts its scheduling strategy according to the changing environment and the type of task. In this research paper we presented a Generalized Priority algorithm for efficient execution of task and comparison with FCFS and Round Robin Scheduling. Algorithm should be tested in cloud Sim toolkit and result shows that it gives better performance compared to other traditional scheduling algorithm.

2.4 Task Scheduling in Cloud Computing

Task scheduling plays a key role in cloud computing systems. Scheduling of tasks cannot be done on the

basis of single criteria but under a lot of rules and regulations that we can term as an agreement between users and providers of cloud. This agreement is nothing but the quality of service that the user wants from the providers. Providing good quality of services to the users according to the agreement is a decisive task for the providers as at the same time there are a large number of tasks running at the provider's side. The task scheduling problem can be viewed as the finding or searching an optimal mapping/assignment of set of subtasks of different tasks over the available set of resources (processors/computer machines) so that we can achieve the desired goals for tasks. In this paper we are performing comparative study of the different algorithms for their suitability, feasibility, adaptability in the context of cloud scenario, after that we try to propose the hybrid approach that can be adopted to enhance the existing platform further. So that it can facilitate cloud-providers to provide better quality of services. Keywords— Cloud Computing, Cloud Architecture, Task Scheduling, Scheduling Types, GA, PSO.

2.5 An Optimized Task Scheduling Algorithm in Cloud Computing:

Cloud provides convenient and on demand network access for computing resources available over internet. Individuals and organizations can access the software and hardware such as network, storage, server and applications which are located remotely easily with the help of Cloud Service. The tasks/jobs submitted to this cloud environment needs to be executed on time using the resources available so as to achieve proper resource utilization, efficiency and lesser makespan which in turn requires efficient task scheduling algorithm for proper task allocation. In this paper, we have introduced an Optimized Task Scheduling Algorithm which adapts the advantages of various other existing algorithms according to the situation while considering the distribution and scalability characteristics of cloud resources.

2.6 Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization:

In cloud computing datacenters exert server unification to enhance the efficiency of resources. Many Vms (virtual machine) are running on each datacenter to utilize the resources efficiently. Most of the time cloud resources are underutilized due to poor scheduling of task (or application) in datacenter. In this paper, we propose a multi-objective task scheduling algorithm for mapping tasks to a Vms in order to improve the throughput of the datacenter and reduce the cost without violating the SLA (Service Level Agreement) for an application in cloud SaaS environment. The proposed algorithm provides an optimal scheduling method. Most of the algorithms schedule tasks based on single criteria (i.e execution time). But in cloud environment it is required to consider various criteria like execution time, cost, bandwidth of user etc. This algorithm is simulated using CloudSim simulator and the result shows better performance and improved throughput.

3. IMPLEMENTATION

The analysis of issues for scheduling in cloud computing was carried out [2] by considering various existing scheduling algorithms like an energy-efficient scheduling application build on private clouds, a scheduling algorithm for private cloud, energy-efficient scheduling of HPC applications in cloud computing environments, an ANT colony algorithm, and workflow scheduling algorithm. The

listed parameters that can be considered for designing a framework for scheduling are 1) Execution time 2) Response time 3) Cost 4) Makespan 5) Scalability 6) Trust 7) Reliability 8) Resource utilization 9) Energy consumption 10) Load balancing 11) fairness.

Disadvantages:

1. Not efficient
2. Time consuming
3. Memory overhead

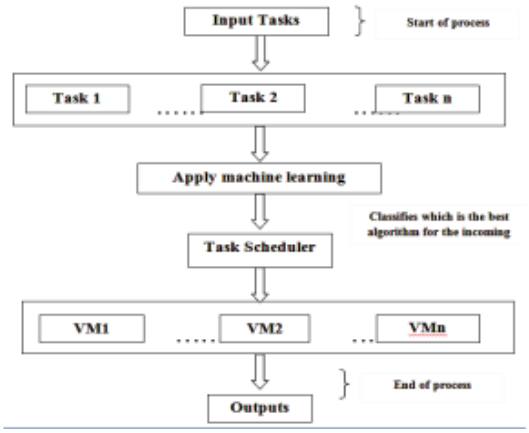

Fig.2 System architecture

In this proposed framework we try to optimize the utilization of cloud computing resources by using machine learning techniques. Task scheduling algorithms can be designed for static or dynamic scenarios. The proposed framework is for the dynamic scenario. Task scheduling can consider different parameters for scheduling purposes like Makespan, QoS, energy consumption, execution time, and load balancing. We propose to apply a machine learning technique for the incoming task requests so as to classify the best suitable algorithm for the task request rather than randomly assigning the scheduling algorithm.

Advantages:

1. The outcome of the proposed work leads to the selection of the best task scheduling algorithm for the input task(request).

In this paper author is using machine learning algorithms such as Logistic Regression and Decision Tree to predict task scheduling resources. All existing algorithms were using random task scheduling or FCFS (first come first server), SJF (shortest job first) and many other algorithms to schedule task in cloud environment but this random scheduling will not assign resources accurately.

To overcome from above problem author is training machine learning algorithms with QOS (quality of service) dataset which contains information about task execution time, make span time, energy and memory consumption. After training machine learning algorithms with above dataset then this trained model applied on new test request to predict resources for that test request and this machine learning algorithms will take all request dynamically and then predict resources. This predicted resources will describe whether task can be currently schedule or has to wait till existing task completed.

4. ALGORITHMS

LOGISTIC REGRESSION:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the

exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc. Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets. Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.

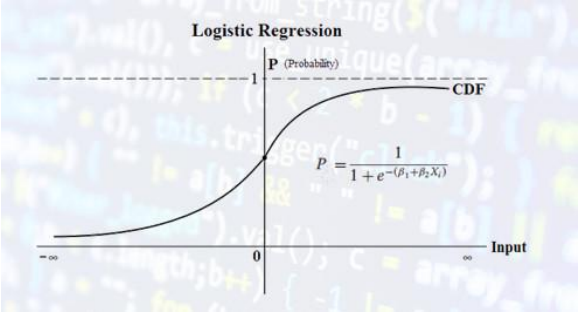

Fig.3: Logistic regression model

DECISION TREE:

Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. We can represent any boolean function on discrete attributes using the decision tree.

Below are some assumptions that we made while using decision tree:

- At the beginning, we consider the whole training set as the root.
- Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
- On the basis of attribute values records are distributed recursively.
- We use statistical methods for ordering attributes as root or the internal node.

In Decision Tree the major challenge is to identification of the attribute for the root node in each level. This process is known as attribute selection. We have two popular attribute selection measures:

1. Information Gain
2. Gini Index

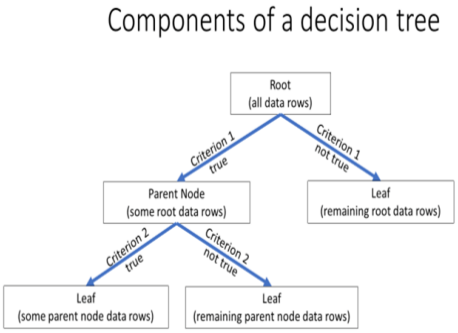

Fig.4: Decision tree model

5. EXPERIMENTAL RESULTS

To trained machine learning algorithms we are using below dataset

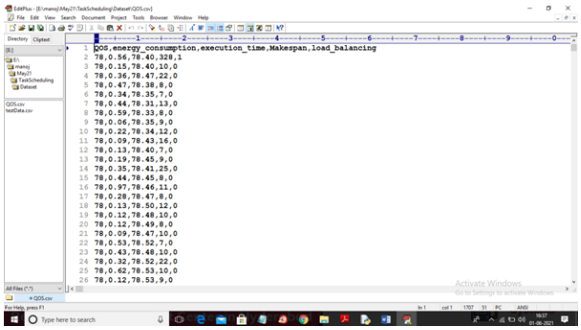

Fig.5: Train dataset

In above dataset in first row you can see dataset column names and remaining rows contains dataset values. In last column we have class label as 0 or 1 where 0 means task can be schedule and 1 means task cannot be schedule at current time. After training with above dataset will use below test data to predict task scheduling.

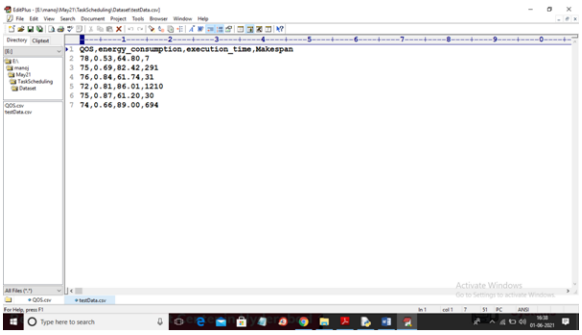

Fig.6: Test data

In above test data we can see there is no class label and machine learning will predict either 0 or 1 as output to indicate scheduling is possible or not.

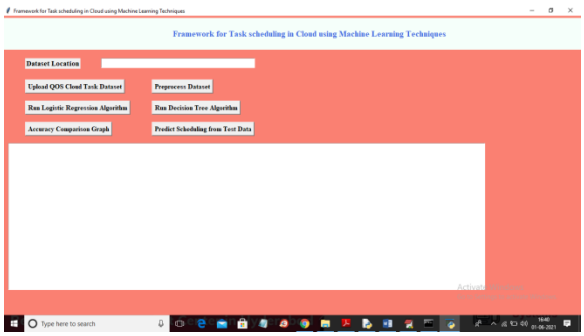

Fig.7: Home screen

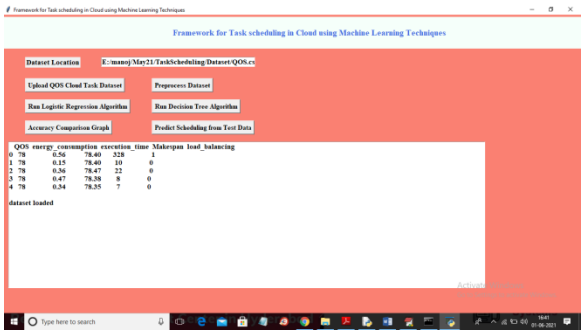

Fig.8: Dataset loaded

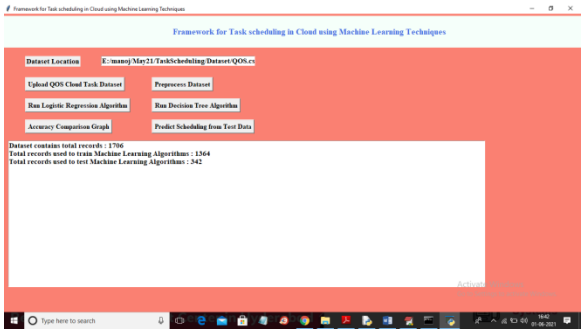

Fig.9: Dataset preprocess

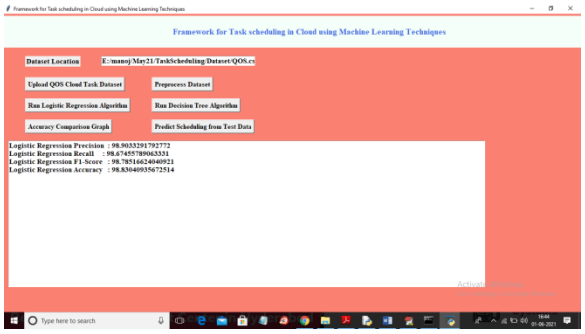

Fig.10: logistic regression algorithm

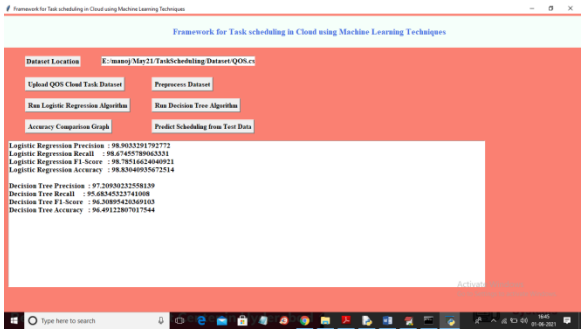

Fig.11: Decision tree algorithm

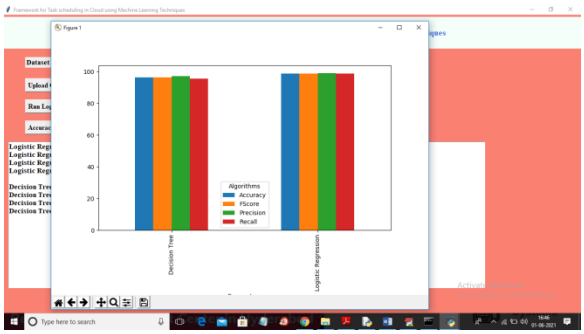

Fig.12:Accuracy comparison graph

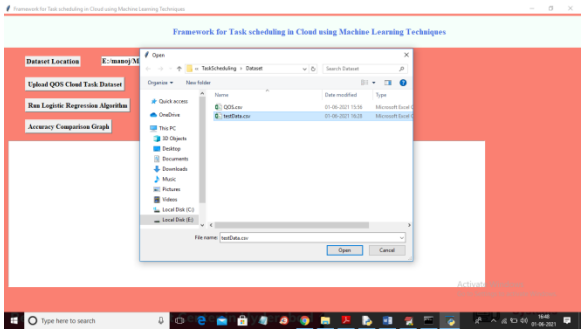

Fig.13: Test data uploading

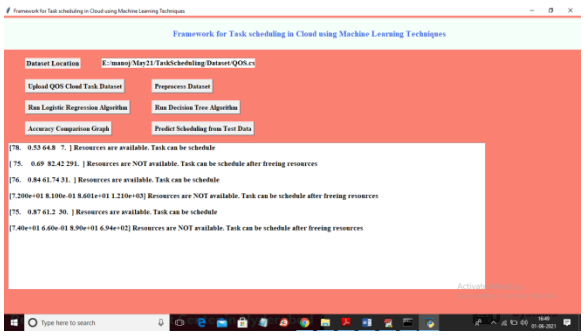

Fig.14: Prediction result

6. CONCLUSION

The selection of appropriate scheduling algorithms for cloud computing is a critical issue. With the correct selection of algorithms, it helps to improve the efficiency of cloud resources. Here we have proposed a method that uses the Machine learning algorithm to classify and appropriate selection of the algorithm. The machine learning used at the application end of the cloud environment gave better results. The proposed method can yield better results compared to traditional random assignment of tasks. We can also compare the traditional method of task scheduling with our proposed method.

REFERENCES

[1]Qi Zhang, Lu Cheng, and Raouf Boutaba ," Cloud computing: stateof-the-art and research challenges", Journal of Internet Services and Applications, Vol.1(1), May 2010.

[2]Vijindra and Sudhir shenai, "Survey on scheduling issues in cloud computing", International conference on modeling optimization and computing (ICMOC-2012), Procidia Engineering 38(2012) 2881-2888.

[3]Dr.Amit Agarwal and Saloni Jain, "Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment", International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 7– Mar 2014 , ISSN: 2231-2803.

[4]Abdul Razaque, Nikhileshwara Reddy Vennapusa, ,Nisargkumar Soni, Guna Sree Janapati and Khilesh Reddy V., "Task Scheduling in Cloud Computing", 2016 IEEE Long Island Systems,Applications and Technology Conference (LISAT).

[5]Shubham Mittal, and Avita Katal," An Optimized Task Scheduling Algorithm in Cloud Computing", IEEE 6th International Conference on Advanced Computing, 2016.

[6] Atul Vikas Lakra, and Dharmendra Kumar Yadav, "Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization",

Procedia Computer Science 48,107 – 113, International Conference on Intelligent Computing, Communication & Convergence, ICCC-2015 .

[7] Mrs.S.Selvarani, and Dr.G.Sudha Sadhasivam, "Improved CostBased Algorithm For Task Scheduling In cloud Computing" , 2010 IEEE International Conference on Computational Intelligence and Computing Research.

[8] Ashwani Kumar Yadav, and Hardwari Lal Mandoria, "Study of Task Scheduling Algorithms in the Cloud Computing Environment: A Review", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 8 (4), 2017, 462-468.

[9] Amanpreet Kaur, and Usvir Kaur, "A Survey for Task Scheduling in Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 5, May 2016 ISSN: 2277 128X.

[10] P. Akilandeswari and H. Srimathi, "Survey and Analysis on Task Scheduling in Cloud Environment", Indian Journal of Science and Technology, Vol 9(37), October 2016.

[11] Ronak Patel, and Hiren Mer, "A Survey Of Various Qos-Based Task Scheduling Algorithm In Cloud Computing Environment", International Journal Of Scientific & Technology Research Volume 2, Issue 11, November 2013 ISSN 2277-8616.

[12] Babur Hayat Malik, Mehwashma Amir, Bilal Mazhar, Shehzad Ali, Rabiya Jalil, and Javaria Khalid, "Comparison of Task Scheduling Algorithms in Cloud Environment", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5, 2018

[13] Hend Gamal El Din Hassan Ali , Imane Aly Saroit, and Amira Mohamed Kotb, "Grouped tasks scheduling algorithm based on QoS in cloud computing network", Egyptian Informatics Journal (2017) 18, 11– 19.

[14] Leila Ismaila, and Abbas Fardoun , "EATS: Energy-Aware Tasks Scheduling in Cloud Computing Systems", The 6th International Conference on Sustainable Energy Information Technology (SEIT 2016), Procedia Computer Science 83(2016) 870 – 877.

[15] Shakeel Ahmad , Imtiyaj Ahmad, and Sourav Mirdha, "A Novel Dynamic Priority Based Job Scheduling Approach for Cloud Environment", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 06 | June - 2017.