EFFICIENT TASK SCHEDULING AND LOAD BALANCING TECHNIQUES TO OPTIMIZE RESOURCE UTILIZATION IN CLOUD ENVIRONMENT

Dr. C.R. Durga Devi Assistant professor, Department of Information Technology, NGM College, Pollachi, India : durgadevi@ngmc.org
Dr.R.Deepa Assistant professor, Department of Computer Science, NGM College, Pollachi,

Abstract-Big data is the collection of data that is huge in size. Big data analysis is the process of analyzing the data and extracting valuable information that can be utilized for future purposes. With the rapid increase in the volume of data, the handling of big data in multiple datacenters with available resources remains a challenging job due to some issues such as higher traffic level, bandwidth, memory and time consumption. The main idea of the proposed Resource Optimized Traffic Aware Gradient Boosting Classification (ROTAGBC) technique is to lessen the traffic occurrence level while distributing the big data in the cloud. The main contribution of the proposed Multivariate Logistic Regression based Gradient Descent firefly optimized Round Robin Scheduling (MLR-GDFORRS) technique is to perform the task scheduling with optimized resource utilization virtual machine at a less time. The main goal of the proposed Support Vector Regression based MapReduce Throttled Load Balancing (SVR-MTLB) technique is to provide better results for performing load balancing with maximum throughput and minimum makespan.
Keywords: Big Data, Task Scheduling, Multivariate Logistic Regression Analysis, Support vector Regression, MapReduce

I. Introduction

Cloud computing is the contemporary type of internet-based computing system. Cloud computing offers a different number of hosted computing services such as storage, servers, computer networks, software, database, and services over the internet to cloud users in an easy manner. By getting these cloud services, many cloud applications are performed in a cloud environment to resolve user requirements. Besides, cloud computing has better advantages such as cost reduction, shared configurable computing resources, computer maintenance, and flexibility in services than the other networks. That is the reason for cloud computing attains great attention among users. Now a days, cloud computing is one of the fastest-growing attractive technology for storing and accessing data from anywhere through the connection of cloud applications over the internet. Cloud computing operates and maintains the different tasks and applications by varying software, platform, infrastructure, and standardizing cloud service providers. Cloud computing is the on-demand network access that plays a key role in sharing the standard computing services where the available resources are delivered as a service over the internet. In cloud computing, valuable information is stored on a data center and accessed by the cloud users with the help of cloud services provided by the cloud service providers. During this, the cloud service providers take full responsibility for organizing and controlling the cloud database with the performance of computing services. In the cloud, the datacenter is the individual server room that is hosted in the association. In order to update the datacenters, the cloud service providers are organized in a cloud environment. Due to the essential factors of the cloud such as scalability, self-service provision, elasticity, pay-per-use, data storage, data recovery, and avoiding attacks, cloud computing is formed by the connection of several servers in individuals and businesses for performing the applications from anywhere.

II. Literature Survey

Sara del Rio et al. (2014) considered a Random Forest classifier to handle imbalanced datasets in big data. Oversampling, under sampling, and cost-sensitive learning were applied to big data with the aid of Map Reduce for handling huge datasets to correctly recognize the under represented class. A novel distributed partitioning methodology was developed by Isaac Triguero et al. (2015) for prototype reduction approaches in nearest neighbor classification. The developed method assisted in characterizing the original training data sets into a minimum number of instances. This in turn aided in improving the speed of the classification process. The storage necessities were minimized with sensitivity to the noise of the nearest neighbor rule. Followed by it, a Map Reduce-based framework was implemented for allocating the functioning of the partitioning method by a cluster of computing elements to combine multiple partial solutions. The developed partitioning method also enabled prototype reduction methods to perform big data classification without accuracy loss. However, the response time was not minimized to the desired level. A linguistic fuzzy rule-based classification system named as Chi-FRBCS-Big Data algorithm was framed by Sara del Rio et al. (2015) for resolving big data classification issues.

A Map Reduce framework was employed with various fusion processes to discover and combine rule bases. Then, an interpretable model was achieved to manage big collections of data with better accuracy and fast response times. This assisted in allocating the computation with the aid of map function and the outputs were integrated by reducing function. But, the Chi-FRBCS-Big Data algorithm failed in improving the classification accuracy. A linguistic cost-sensitive fuzzy rule-based classification method referred to as the Chi-FRBCS-BigData algorithm was designed by Victoria Lopez et al. (2015) for handling imbalanced big data. A fuzzy rule-based classification system was established to handle uncertainty in large volumes of data. Then, the Map Reduce framework was employed to allocate the computational operations of the fuzzy model. Followed by it, cost-sensitive learning approaches were also employed to address the data imbalance. Though computation time was reduced, the Chi- FRBCS-Big Data CS algorithm failed in considering the processing time.

An open-source, distributed framework named Inter IMAGE Cloud Platform (ICP) was considered by V. A. Ayma et al. (2015) for automatic image interpretation. ICP enabled Data Mining Package for executing supervised classification procedures on large amounts of data with the aid of Hadoop Map Reduce. Classification methods were also implemented. An SVM classifier was applied to data sets of various sizes for attaining different cluster configurations. But, Inter IMAGE failed in the efficient scheduling of tasks. SamipRaut et al. (2016) described an algorithm for minimizing the overall data allocation and processing time. The overall data processing time in the cloud was minimized efficiently. Although the processing time was reduced, secure big data storage was not considered. Jianguo Chen et al. (2017) described Parallel Random Forest (PRF) algorithm for big data. Depended on a hybrid approach combining data-parallel and task-parallel optimization, the PRF algorithm was optimized. Then, a vertical data-partitioning approach was carried out for minimizing the data communication cost. Followed by it, a data-multiplexing approach was employed to reuse the training dataset and lessen the data volume. A dual parallel method was also introduced in the training process of RF. Subsequently, a task Directed Acyclic Graph (DAG) was generated concerning the parallel training process of PRF and reliance on Resilient Distributed Datasets (RDD) objects. Different task schedulers were appealed for the tasks in DAG. In addition, the dimension-reduction method in the training process and the weighted voting method in the prediction process were applied for enhancing the accuracy of the PRF algorithm for huge, high-dimensional data. But, energy-saving remains unaddressed.

III. Methodology

The proposed system is focused on efficient task scheduling and load balancing performance with higher efficiency and minimum time consumption. The proposed research work is designed with three different techniques such as Resource Optimized Traffic Aware Gradient Boosting Classification (ROTAGBC) technique, Multivariate Logistic Regression based Gradient Descent firefly optimized Round Robin Scheduling (MLR-GDFORRS) technique, and Support Vector Regression based Map Reduce Throttled Load Balancing (SVR- MTLB) technique.Figure 1.1 shows the framework of the Proposed System. The proposed ROTAGBC technique is implemented with the aim of performing the traffic aware big data distribution with optimized resource utilization in the cloud. Theproposed ROTAGBC technique uses prioritizing the incoming tasks by using job request parameters. Gradient boost ensemble classifier is used to categorize the number of tasks as immediate or reserved. This helps to improve the classification accuracy. The users send the number of incoming tasks. The proposed technique schedules the incoming tasks to VMs and it completes the job in parallel on a host with various finishing times. The task is referred to as the processing of multiple user requests. The priority is assigned by the task assigner to the incoming tasks.

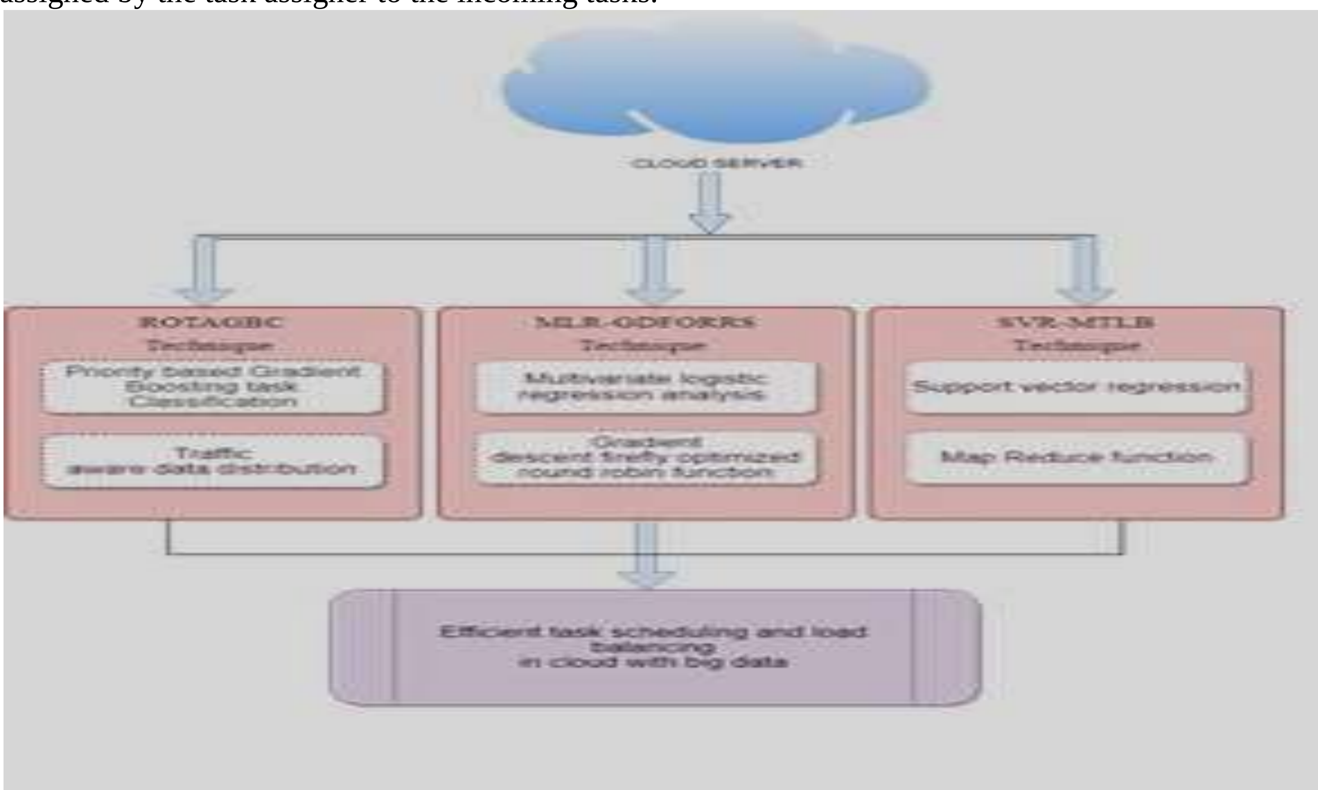

Figure 1.1 Framework of the Proposed System

Based on the job request parameters such as task size, bandwidth, and memory capacity. According to the priority level, the tasks are categorized with the gradient boosting classification technique. Then, the prediction model is applied as an ensemble of weak decision trees by resource optimized gradient boosting classifier. Followed by, the Gradient boost ensemble classifier is used to categorize the number of tasks as immediate or reserved for assigning the task to multiple datacentres in the cloud. Every base classifier is merged to construct the strong classifier for achieving maximum efficiency. Next, the categorized tasks are stored during task assigner and distributed over multiple datacenters in various positions with optimal resource utilization. Therefore, the ROTAGBC technique minimizes the traffic occurrence level in the big data distribution. The proposed MLR-GDFORRS technique is introduced with the objective of minimizing the time consumption and workload to execute the task scheduling via an optimal virtual machine. The objective is achieved with the help of designing the multivariate logistic regression analysis and gradient descent firefly optimized round robin scheduling approach. The cloud server gathers the number of tasks from dissimilar locations and stores the user request depending on their priority. The cloud manager uses the multivariate logistic regression analysis for discovering the relationship amongst dependent data and one or more independent data of similar cloud user requests by observing the tasks. The priority level is changed while some changes occurred in the job request parameters value. The regression analysis is used to schedule the tasks for the virtual machine by the cloud manager. Next, the tasks are stored in one or more queues according to the priority level for reducing the space complexity.

After that, the Gradient descent firefly optimized Round Robin scheduling is employed to assign the tasks to the resource optimized virtual machine after choosing the resource optimized virtual machine. By using firefly optimized algorithm, the resource optimized virtual machine is discovered between the numbers of the virtual machine based on the light intensity. Based on the light intensity, weightage and rank is assigned to the virtual machines. The cloud manager selects the virtual machine with high rank for scheduling the tasks. The minimum resource utilization of the virtual machine is chosen for managing the maximum priority task. Therefore, traffic level and the probability of data loss are reduced. Then, the Map Reduce function is applied to allocate the maximum priority tasks to the optimal virtual machine with minimum time consumption. This helps to reduce the workload of the cloud server in an effective manner. In addition, the MLR-GDFORRS technique offers better results for big data analysis in the cloud with higher task scheduling efficiency.

The proposed SVR-MTLB technique is implemented with the application of the support vector regression analysis and Map Reduce function to enhance the load balancing performance between data centers. The SVR-MTLB technique aims to maximize resource utilization rate and minimum time consumption in big data. Also, the workload of the virtual machine is reduced for distributing the tasks over multiple data centers in the cloud. Cloud users send a huge number of requests (data/files) to cloud servers in dissimilar locations. After transmitting cloud user requests, the support vector regression is utilized to balance the workload of the virtual machine. The support vector regression is carried out to identify the current resource status of the virtual machine in the index table and measure the capacity, bandwidth, and memory of the virtual machine. The load balancer applies the index table for managing the virtual machines. Therefore, the migration time is reduced. Then, the regression analysis is performed by using the map function optimal hyperplane. The map function is offered the three resource statuses of the virtual machine namely overloaded, less loaded, and a balanced load. Followed by, the hyperplane is applied as a boundary with aid of support vector regression for discovering the overloaded and less loaded virtual machine. With this determination, the ID of a less loaded virtual machine is sent by the load balancer to the data center controller. The tasks are migrated from an overloaded VM to a less loaded VM at run time with the help of the controller. The MapReduce function is utilized to perform the migration of tasks to find the resource capabilities of the virtual machine. In addition, it is used to forecast the less loaded virtual machine. This helps to reduce the response time. The three proposed techniques namely the ROTAGBC technique, MLR-GDFORRS technique, and SVR-MTLB technique are implemented in Java language.

IV.EXPERIMENTAL EVALUATION

The assessment of the proposed method is carried out for efficient task scheduling in the cloud based on the metrics such as Scheduling Efficiency (SE), Scheduling Time (ST), Resource utilization rate (RUR), Space Complexity (SC), and Error Rate (ER). The performance of the proposed MLR-GDFORRS technique is compared with the existing HGDCS and IPSO algorithms in terms of parameters Scheduling Efficiency, Space Complexity, True Positive Rate, Scheduling Time, and Resource Utilization Rate. Table 7.19 shows the improvement values in terms of percentage as applied for the Amazon EC2 Dataset and the Personal Cloud Dataset.

Table 1.1: Overall Performance Improvement in Task Scheduling

TASK SCHEDULING PARAMETERS	**Amazon EC2 Dataset**		**Personal Cloud Dataset**	
	IMPROVEMENT (%)		**IMPROVEMENT (%)**	
	Existing HGDCS vs Proposed MLR-GDFORRS	Existing IPSO vs Proposed MLR-GDFORRS	Existing HGDCS vs Proposed MLR-GDFORRS	Existing IPSO vs Proposed MLR-GDFORRS
Scheduling Efficiency	**18.7962**	**24.5681**	**16.7002**	**19.0672**
Resource Utilization Rate	**21.9227**	**34.7664**	**21.3489**	**32.1056**
True Positive Rate	**27.2806**	**14.0474**	**20.7048**	**14.2533**
Space Complexity	**24.6777**	**13.3135**	**27.4792**	**14.9373**
Scheduling Time	**30.1254**	**15.4994**	**23.1153**	**16.6066**

From table 1.1, it is clear that the Scheduling Efficiency attained an improvement of 18.79% Resource Utilization Rate 21.92%, True Positive Rate obtains an improvement of 27.28%, Space Complexity obtains 24.67% and the Scheduling Time obtains 30.12% of improvement as compared with existing HGDCS algorithm using the Amazon EC2 dataset. Scheduling Efficiency attained an improvement of 24.56% Resource Utilization Rate 34.76%, True Positive Rate obtains an improvement of 14.04%, Space Complexity obtains 13.31% and the Scheduling Time obtains 15.49% of improvement as compared with the existing IPSO algorithm using Amazon EC2dataset.

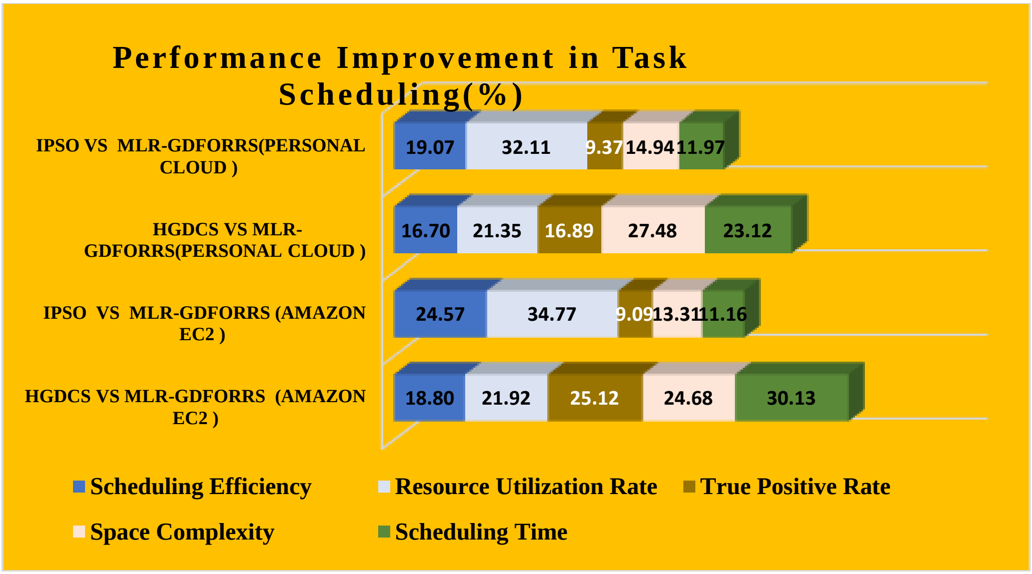

Figure 1.2: Measurement of Performance Improvement in Task Scheduling

The Scheduling Efficiency attained an improvement of 16.70% Resource Utilization Rate 21.34%, True Positive Rate obtains an improvement of 20.70%, Space Complexity obtains 27.47% and the Scheduling Time obtains 23.11% of improvement as compared with the existing HGDCS algorithm using the Personal Cloud dataset. Scheduling Efficiency attained an improvement of 19.06% Resource Utilization Rate 32.10%, True Positive Rate obtains an improvement of 14.25%, Space Complexity obtains 14.93% and the Scheduling Time obtains 16.60% of improvement as compared with the existing IPSO algorithm using the Personal Cloud dataset.Figure 1.2 shows the graphical representation of the improvement of Task Scheduling using the MLR-GDFORRS technique.

V.CONCLUSION

In our research work, three proposed techniques, namely ROTAGBC, MLR-GDFORRS, and SVR-MTLB, have been introduced aiming at enhancing the performance of resource optimized traffic-aware task scheduling with maximum throughput and minimum makespan. With the objective of performing the big data distribution over multiple datacenters with less traffic occurrence levels, the proposed Resource Optimized Traffic Aware Gradient Boosting Classification (ROTAGBC) technique is introduced in the first research work. In the second work, the proposed Multivariate Logistic Regression based Gradient Descent firefly optimized Round Robin Scheduling (MLR-GDFORRS) technique is introduced aims at performing the large task scheduling to the optimal virtual machine with less amount of time consumption and less workload during the big data distribution in cloud. The proposed Support Vector Regression based MapReduce Throttled Load Balancing (SVR-MTLB) technique is introduced for providing better results in the load balancing with higher throughput and less makespan

REFERENCES

[1]. Fan-Hsun Tseng Xiaojiao Chen, Li-Der Chou, Han-Chieh Chao and Shiping Chen “Support vector machine approach for virtual machine migration in cloud data center”, Multimedia Tools and Applications, Springer, Volume 74, Issue 10, pp. 3419-3440, May 2015.

[2]. Heba Saleh, HebaNashaat, Walaa Saber, and Hany M. Harb, “IPSO Task Scheduling Algorithm for Large Scale Data in Cloud Computing Environment”, IEEE Access, Volume 7, pp. 5412-5420, March 2019

[3]. Kaibin Li, Zhiping Peng, Delong CuiandQirui Li, “SLA-DQTS: SLA Constrained Adaptive Online Task Scheduling Based on DDQN in Cloud Computing”, Applied Science, MDPI, Volume 11, Issue 20, pp. 1-17, 2021

[4]. Mikel Elkano, Mikel Galar, JoseSanz, and HumbertoBustince, “CHI-PG: A fast prototype generation algorithm for Big Data classification problems”, Neuro computing, Elsevier, Volume 287, pp. 22-33, 2018.

[5] Bela Shrimali and Hiren Patel, “Multi-objective optimization oriented policy for performance and energy efficient resource allocation in Cloud environment”, Journal of King Saud University - Computer and Information Sciences, Elsevier, December 2017, Pages 1-10.

[6] Ibrahim AbakerTargio Hashem, Nor BadrulAnuar, Mohsen Marjani, Abdullah Gani, Arun Kumar Sangaiah, AdewoleKayodeSakariyah, “Multi-objective scheduling of MapReduce jobs in big data processing”, Multimedia Tools and Applications, Springer, Volume 77, Issue 8, 2018, Pages 9979–9994.

[7] Mikel Elkano, Mikel Galar, JoseSanz and HumbertoBustince, “CHI-PG: A fast prototype generation algorithm for Big Data classification problems”, Neuro computing, Elsevier, Volume 287, 2018, Pages 22-33.

[8]. Syed Hamid Hussain Madni, Muhammad ShafieAbdLatiff, Shafi’i Muhammad Abdulhamid and Javed Ali, “Hybrid gradient descent cuckoo search (HGDCS) algorithm for resource scheduling in IaaS cloud computing environment”, Cluster Computing, Springer, Volume 22, October 2019, Pages 301 – 334.