A Novel Approach for Transaction Management in Heterogeneous Distributed Real-Time Replicated Database Systems

Mr. Dheeraj Bhimrao Lokhande[1], Prof. Dr. P. K. Deshmukh[2]
[1]Research Scholar, P. A. H. Solapur University, Solapur.
[2]Professor & Head, Shriram Institute of Engg. Solapur,

Abstract: Today, Cloud computing is most widely used system due to number of benefits to end users. In IT (Information Technology) organizations, cloud computing is most vital domain to work. There are different types of services such as SaaS, IaaS and PaaS provided by cloud computing based on the end users needs. Rather than using own resources for data storage and Management, organizations started to utilize cloud computing data storage systems. Cloud storage (eg. Amazon S3) is emerging as a popular service and most of the enterprises shifts their data workloads to the cloud. The popularity of Cloud computing system increased dramatically because it rent computing resources, bill on a pay-as-you-go basis and multiplex many users on same physical infrastructure. Cloud users are provided an illusion of infinite computing resources so that on demand resource consumption rate can be increased or decreased.

Cloud computing offers the vision of a virtually infinite pool of computing, storage and networking resources where applications can be deployed. It is the popular solution for on-demand and dynamic resource provisioning. The provisioning & maintenance of cloud resources are done with the help of resource management techniques(RM). The RM techniques are responsible to keep the track of free resources and assign the resources from free pool to incoming tasks. Along with the growing demands of modern applications and workloads, cloud computing has gained prominence throughout the IT industry. Modern web applications are growing along different dimensions such as the number of users, complexity, data size and many people are connecting to the internet through various devices.

Modern web applications requires more performance and capacity from their data storage and management systems, so database systems are normally used. As web applications become more popular, it can experience exponential growth from users all over the world. So, to reduce the response times many web applications use data centers in order to allow users to access the data center. Web Applications which deals with the cloud storage systems contain heterogeneous workload to deploy on the cloud. This heterogeneous workload needs to be handled carefully otherwise problems like long scheduling delay, starvation for low priority tasks may occur which can hurt the performance of the application significantly. To handle this workload characterization is much more important. In workload characterization, heterogeneous workload is divided into multiple task classes with similar characteristics in terms of resource and performance objectives. It is important to consider heterogeneity of workload. The workload typically consists of diverse applications with different priorities and resource requirements. Failure to consider heterogeneous workload, it will lead to long scheduling delay, starvation by affecting performance of the application. Modern applications faces challenges such as workload characterization, resource allocation and security. To meet these needs we propose a new framework which interacts with the heterogeneous database systems and provides workload characterization by using K-means clustering algorithm, resource allocation by providing virtual machines through requesting to the cloud and security by using AES algorithm for encryption and decryption.

In this research work, transactions are taken in form of heterogeneous files from users as input and deploy the files efficiently at cloud storage system and on heterogeneous database systems as output. Transactional operations like insert, update, delete are performed on users data at cloud storage systems and on heterogeneous database systems. The transactional query operations are performed on database systems. Here, the proposed system attempts to design novel approach for secure, energy efficiency, scalable transaction management system based on terminologies and methods of resource scheduling in heterogeneous distributed database systems.

The thesis addresses "the transaction management problem", which is to interact with the heterogeneous database systems and to allocate and schedule computing resources with workload characterization by providing security to user's data.

Performance Evaluation of the proposed system is done based on fluctuating workloads and in terms of Throughput, Energy Efficiency, CPU Utilization, Scheduling Time, Response Time. The performance of the system is measured and compared with Container Based Scheduling(CBS), Dynamic Power–Saving Resource allocation(DPRA) mechanism and SLA-Aware Energy resource management(SLA) technique.

The Experimental results of the proposed system shows effective performance of the system. Results shows that proposed system outperforms existing techniques. The implemented environment is easy to use and the input to the system are given for execution in a normal way without changing or restructuring the code.

Keywords: Cloud Computing, Cloud Storage, Workload Characterization, Energy Efficiency, Resource Management, Resource Allocation.

I. INTRODUCTION

In this research work, after analyzing the heterogeneous workload task classification is performed by using k-means clustering algorithm. Using standard K-means clustering, we show that the heterogeneous workload can be divided into multiple task classes with similar characteristics in terms of resource and performance objectives. After workload analysis it can be found that workload composition are highly heterogeneous and dynamic over time.In this research work, heterogeneity of the workload is considered. Specifically, we first use the K-means clustering algorithm to divide workload into distinct task classes with similar characteristics in terms of resource and performance requirements.The proposed framework which is designed considers the heterogeneous nature of workload and performs following three operations:-

A. Dividing tasks into task classes using the K-means algorithm. whose main objective is to understand workload characteristics
B. Once workload characterization is obtained then allocate the resource i.e. Virtual machines to the characterized task.
C. Perform security operations on the task before uploading on cloud storage systems.

To explain the framework, Amazon AWS cloud is used. The work mainly focuses on workload characterization using clustering of the input workload, allocate the resource by using AWS EC2 and performing security operations i.e. encryption and decryption on assigned work. The basic model of the framework is shown in the following figure.
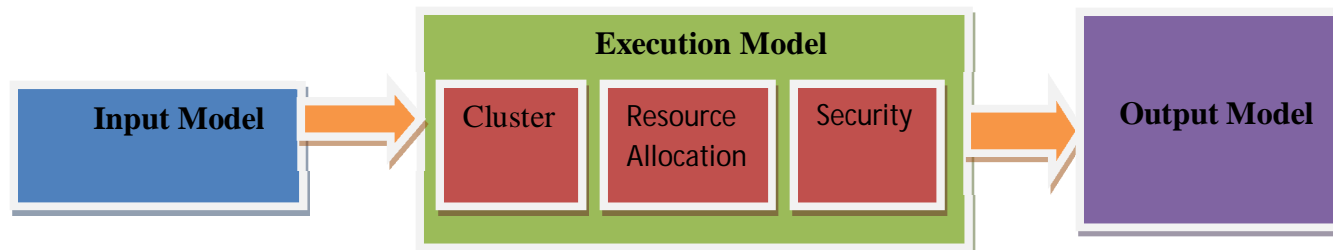

Figure :- Basic Model of Proposed framework

The proposed framework contains three important models:-
1) Input model
2) Execution model
a) Clustering
b) Resource Allocation
c) Security i.e. Encryption and Decryption
3) Output model

The framework communicates with Amazon AWS cloud to use the various services provided by the Amazon. The services are:-
a) *Elastic Compute cloud (EC2):* EC2 is used to allocate the resources i.e. Virtual Machines
b) *DynamoDB:* DynamoDB is used to create cloud database which stores important information regarding users and files.
c) *Simple Storage System (S3):* The framework performs encryption operation on the files and these encrypted files are stored on S3. So for storage S3 is used where a special space is created for framework used called buffer.

The framework utilizes MySQL database to store user and file related information. In MySQL, 'transaction' database is created which is collection of various tables.The framework uses k-means clustering algorithm for workload characterization and for establishing task classes. These classes are formed based on type of the file, size of the file etc. Based on that two clusters are created. These clusters are allocated resources i.e. virtual machines by using Amazon EC2 cloud.

II. EXPERIMENTAL SETUP

The framework is implemented in Netbeans IDE 8.2 environment. It contains different files for clustering, resource allocation and security. The framework asks users about registration details. All this information regarding user is kept in both the databases i.e. MySQL database and in Amazon AWS DynamoDB database. So MySQL server 5.1 is installed by using mysql-essential-5.1.54-win32 setup file. To use MySQL database with graphical user interface (GUI) SQLyog community 13.1.1 -32 bit is used.The Amazon AWS Cloud is connected to the framework by using credential file which includes aws_access_key_id and aws_secret_access_key i.e access id and security key to access AWS cloud. The performance of the framework is measured on computer system of core i3 supporting NetBeans environment with windows 10 operating system. The input files to the framework is taken from user specified path. The input files are of heterogeneous type with includes .doc files,.ppt files,.txt files, image and video files. The framework takes the input files in the range 01,05,10,15,20,25,30.The framework clusters these files, allocate resources and perform encryption and decryption on users data.

A. Steps of execution

The step wise execution procedure of the system scenario is shown pictorially. Following few figures are depicting the actual working steps and procedure of transaction management framework. Before executing the framework, first MySQL database is started by using SQLyog Community which GUI tool for RDBMS MySQL database. In MySQL, 'transaction' database is created which is collection of table such as registration table, encryption table etc.At the same time, we have started Amazon AWS Cloud by using user name and password. The following services of Amazon AWS cloud is used by the framework.

1) *EC2 (Elastic Compute Cloud):* This service provides resources i.e. virtual machines required by the framework. Dynamically the framework allocate the resources (VM) to user. The framework requests twice for the virtual machines so that EC2 allocate two VM's.
2) *DynamoDB:* This service is provided by Amazon AWS cloud to create database. It is the cloud database used by the framework. It is a collection of tables such as user registration table, keystore table, file table etc.
3) *S3 (Simple Storage Service):* This service is provided by Amazon AWS cloud for Storage. Huge amount of data can be stored on cloud storage systems such as S3.In this, a special buffer is created for the framework users named 'Transaction2122'.In this buffer, for every user separate space is allocated. The name of folder is given on the email id of every user.

B. Performance Evaluation

The framework is implemented on Netbeans 8.2. The framework utilizes various service provided by the Amazon AWS Cloud. The framework uses another system with different configuration for replication. Another system consist of 'MongoDB' database which is used by the framework to store replicated values. 'MongoDB' is a NoSQL database.The framework shows the performance in graphical form after the 'Performance' button is clicked by user. All the values generated are shown in graphical form by the framework. The performance of the framework is evaluated by using several parameters.

The parameters are as follows:-

1) Throughput
2) Datarate
3) Time to Encrypt
4) Time required with Encryption and without Encryption
5) CPU Utilization
6) Time to Upload
7) Delay Time

a) *Throughput:* It is a measure of how much amount of time is required to upload an average of files on the cloud storage systems.

It is calculated as:-

$$\text{Average value} = \frac{\text{Total File Size}}{\text{Number of Files}}$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

b) CPU Utilization: It is a measure of time which shows how much percentage of CPU is used by the framework while uploading a file. It can be calculated as:-

$$CPU\ Utilization = Usages\ (\ ).\ cpuuti;\quad Where$$

The variable cpuuti is used find out the load on CPU.

c) Delay Time: It is the total time taken by the framework from start of the uploading process to end of the uploading process. The method named current Time Millis () is used to calculate the delay time. It is calculated as:-

$$Delay\ Time\ =\ \frac{(\ End\ Time - Start\ Time)}{1000}$$

d) Datarate: Datarate graph shows in a millisecond how much file data uploaded on the cloud storage systems. It is calculated as:-

$$FSize\ =\ \frac{File\ Size}{1000}$$

$$Datarate\ =\ \frac{FSize}{Delay\ Time}$$

e) Upload Time: The upload time is the time required to upload a file. The value is calculated in milliseconds. The method call current Time Millis ( ) used to find the required time. It can be calculated as:-

$$Uploading\ Time\ =\ \frac{(\ Stop\ upload - Start\ upload\ )}{1000}$$

f) Encryption Time: It is the taken by the framework to encrypt a single file. The value is calculated in milliseconds. Encryption Time is calculated as:-

$$Encryption\ Time\ =\ \frac{[(EncEnd - EncStart)]}{1000}$$

III. EXPERIMENTAL METHODOLOGY

The framework is implemented completely in Java Netbeans IDE environment. The experiment is conducted on a system with core i3 machine on the windows10 platform. Connection with Amazon AWS cloud is made by using user Id and password. Then various services such as EC2, DynamoDB, S3 is set to be used by the framework to use and store necessary information in appropriate place.At the same time MySQL database is used by using SQLyog Community Software. The SQLyog Community contains necessary information about the framework. It is a collection of tables which contains important information regarding the framework such as user registration details, generated key's for the files, files in the encrypted formats etc.The input files given to the framework are varied in range 01, 05, 10, 15, 20, 25 and 30.According clustering is performed on user input and cluster are created by using file type, file size. These clusters are allocated to virtual machines for uploading. By the request through coding i.e. dynamically, Amazon AWS EC2 service, provides the two virtual machines required for the framework to store the files on Amazon S3 Cloud. The database named 'DynamoDB' which is present on the cloud contains number of table to store important information regarding the files. It contains information like uploaded files, Encrypted files, different keys of different files, information about deleted files in the various tables.The storage service of Amazon S3 (Simple Storage System) cloud used to create separate buffer for the framework. Each user on the framework is allocated separate space to store their files. The framework takes number of files from user as input then encrypts the files and store the files on AWS S3 storage on appropriate user's folder or space. Administrator named 'PKG' is created by the framework which identifies the authenticated users on network. Only after Administrator permission, each user is able to work with the files on the network and user can download file, update the file or delete the file. The user can not delete the file directly on network. The delete request is first sent to Administrator. The administrator verifies the authenticity of the

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

user and then delete permission is given to user. The performance of the framework is checked by using parameters like throughput, CPU utilization, Energy Efficiency, Scheduling Time and Response Time. The framework is compared with Container Based Scheduling Technique (CBS), Dynamic Power Saving Resource Allocation (DPRA) method and SLA-Aware Energy Efficient Resource Management for Cloud Environments (SLA) method.

IV.   METHOD TO CALCULATE ENERGY EFFICIENCY

In this research work, encryption operation is performed on the user specified files while uploading these files on cloud storage systems. The framework encrypts the files as the uploading options chosen by the user such as 01, 05, 10, 15, 20, 25 or 30 files. It is necessary to encrypt these files to protect data from unauthorized access.

The framework notes the number of runs for different number of files and the battery life. The difference in battery life left is noted for uploading files. Change in battery life divided by the number of files gives the battery life consumed in percentage.

Consider the following table which contains number of files which are encrypted before uploading on cloud storage systems and percentage of battery left after performing encryption operation. Battery performance is measured when percentage of battery left 99 %.

Number of Files	Percentage Battery Left (%)
01	98
05	96
10	94
15	92
20	90
25	88
30	86

Table :- Sample table for energy efficiency calculation

As seen in the above table, for a file to encrypt the battery percentage reduced from 99 to 98 percentage. so the energy consumption for a file is one percent (01%).Similarly, for 05 files the battery is change or reduced from 98 to 96 percentage so energy consumption rate is two percent (2%).In this way the energy efficiency is calculated which indicated percentage of battery used for encryption operation.

V.   EVALUATION PARAMETERS

The framework deals with heterogeneous workload by clustering the workload, allocating the resources to the clusters and proving security to user data. The framework is compared with Container Based Scheduling(CBS), Dynamic Provisioning Resource Allocation(DPRA) and SLA-Aware Energy Efficient Resource Management for Cloud Environments(SLA).The main parameters considered for evaluating the performance of the framework are as follows:-

1) Throughput
2) CPU utilization
3) Energy Efficiency
4) Scheduling Time
5) Response Time

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

VI. RESULT

A. Throughput

The results after executing the files in the range starting from input files 01,05,15,20,25,30 are collected in tabular form for CBS, DPRA, SLA and proposed work .The input files for the framework can be from 1 to 30.so framework can upload 30 files with encryption on cloud storage system. The following table shows throughput values for the CBS, DPRA, SLA and proposed work respectively.

Throughput

No of Files	CBS	DPRA	SLA	Proposed
1	34.6	37.6	30.2	44.9
5	35.6	37.2	28.2	42.9
10	33.8	35.5	31.2	49.6
15	33.1	34.2	25.8	36.6
20	33.5	36.2	25.7	37.9
25	36.7	38.6	33.2	40.4
30	43.6	45.8	38.3	54.6

Table:- 'Throughput' table

The throughput table shows that the values are calculated for files 01,05,10,15,20,25,30 resp.
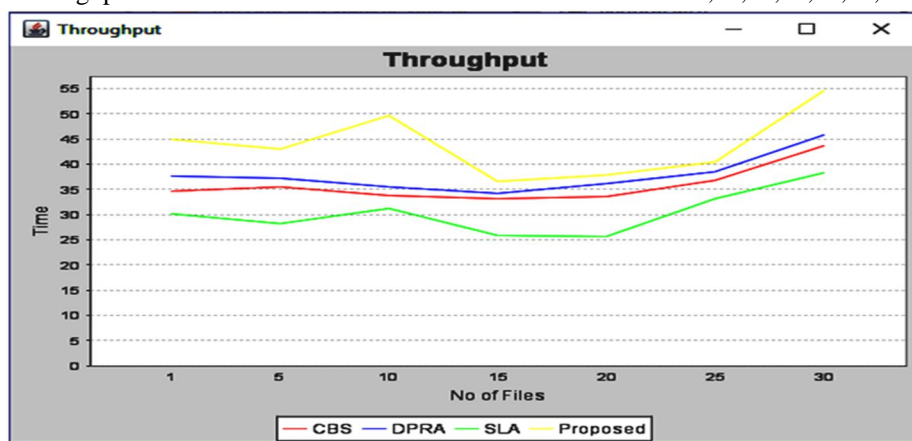

Figure:- 'Throughput' Graph

B. CPU Utilization

It is a measure of how much percentage of CPU is used by the methods to upload the files on cloud storage systems.The following table shows the CPU Utilization values for CBS, DPRA, SLA and Proposed work respectively.

CPU Utilization

No of Files	CBS	DPRA	SLA	Proposed
1	75.5	60.6	84.5	44.5
5	60.5	75.5	84.5	56.5
10	70.3	55.7	68.5	50.2
15	73.3	63.6	82.4	52.3
20	69.2	74.1	58.2	52.2
25	64.1	61.0	73.0	50.1
30	75.3	47.0	75.4	45.4

Table :-' CPU Utilization' table

The CPU Utilization table shows the values are calculated by running files in the order 01,05,10,15,20,25 and 30.The table suggest that proposed work uses CPU much lesser than the other systems.The values shown in the tabular form are shown in graphical form by the following figure.
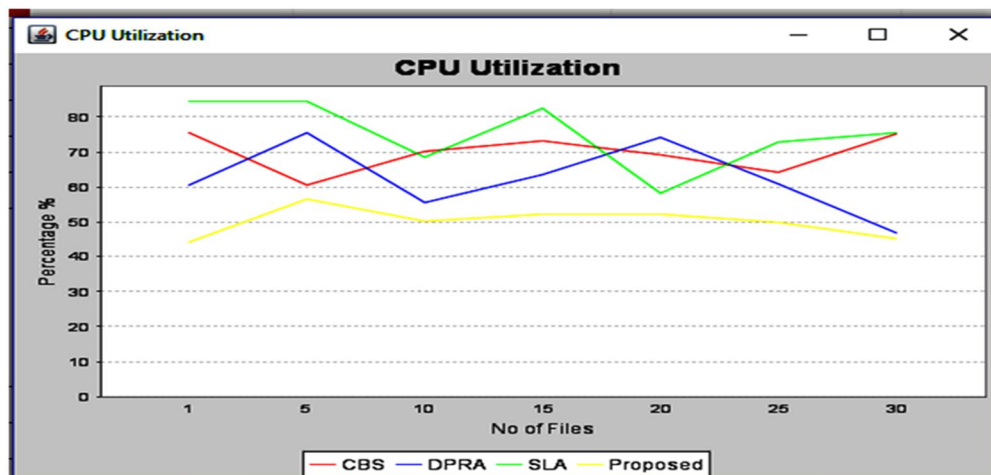
Figure:- ' CPU Utilization' Graph

C. Energy Efficiency

The value is calculated for encryption process. It is a measure of change in battery life divided by the number of files. It gives the battery life consumed in percentage.

Energy Efficiency

No of Files	CBS	DPRA	SLA	Proposed
1	1.0	1.2	1.5	0.5
5	1.2	1.8	2.0	1.0
10	2.0	2.7	2.8	1.5
15	2.8	3.0	3.5	2.5
20	3.2	3.3	3.8	3.0
25	3.6	3.5	4.0	3.1
30	3.7	3.8	4.8	3.2

Table:- 'Energy Efficiency' table

The values for Energy Efficiency are calculated by executing the framework for 01, 05, 10, 15, 20, 25 and 30 files in order. The energy efficiency is calculated for encryption process i.e. how much percentage of battery power is used to encrypt files by the approaches. As it can be seen in the table, the values for energy efficiency for proposed work are less than CBS, DPRA and SLA methods respectively.
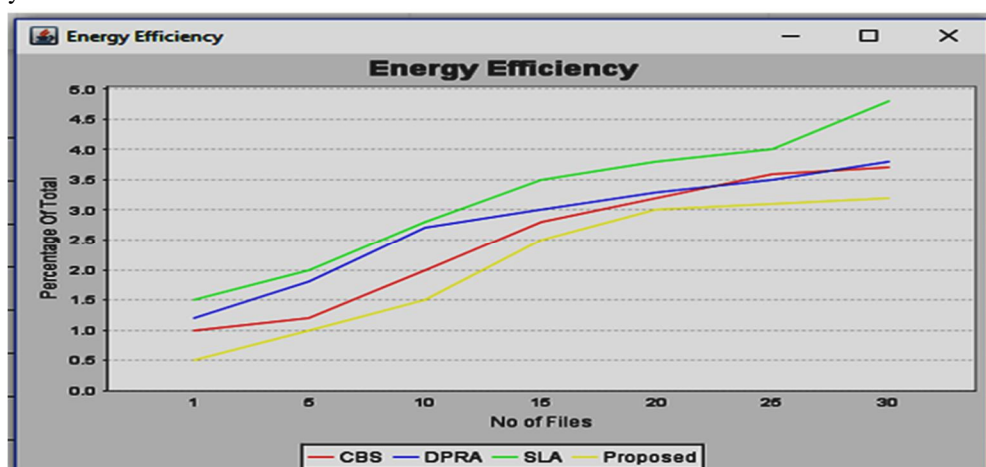

Figure:- 'Energy Efficiency' Graph

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

D. Scheduling Time

Scheduling Time is the time file has to wait to schedule for encryption. It is calculated in milliseconds. The following table shows the Scheduling Time values for CBS, DPRA, SLA and Proposed work respectively.

Scheduling Time

No of Files	CBS	DPRA	SLA	Proposed
1	27.9	17.6	36.3	6.2
5	47.4	37.1	56.2	35.5
10	74.7	65.1	83.3	53.2
15	108.1	98.2	117.3	87.3
20	140.2	130.3	148.4	118.8
25	166.7	158.2	176.1	145.6
30	200.3	190.1	209.5	180.2

Table:- 'Scheduling Time' table

The values in the scheduling Time table is calculated by executing the framework 01 ,05, 10, 15 , 20 ,25 and 30 files one after the other. As the values suggest in the table that the proposed work requires less time than the existing systems i.e CBS ,DPRA and SLA.The following figure shows the graph for scheduling Time.
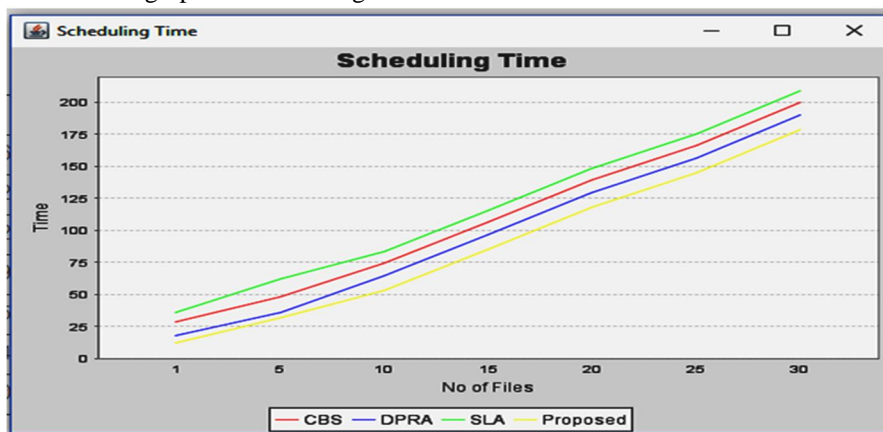

Figure: 'Scheduling Time' Graph

E. Response Time

Response Time is the total time taken by the framework for execution. It is calculated in seconds.
The following table shows Response Time for CBS, DPRA, SLA and Proposed work.

Response Time

No of Files	CBS	DPRA	SLA	Proposed
1	179.8	169.8	189.3	154.7
5	175.4	150.1	165.4	115.3
10	300.4	300.7	285.3	245.4
15	237.5	257.6	297.5	222.4
20	328.6	338.5	299.7	279.5
25	319.1	399.6	320.1	296.1
30	429.6	403.3	441.3	383.6

Table:- 'Response Time' table

The Response Time table is generated by the framework after uploading the file in the order 01, 05, 10, 15, 20, 25 and 30. As the values in the table indicated that the proposed work requires less time than the other existing approaches i.e. CBS , DPRA and SLA approach.The related graph of the Response table is presented below.
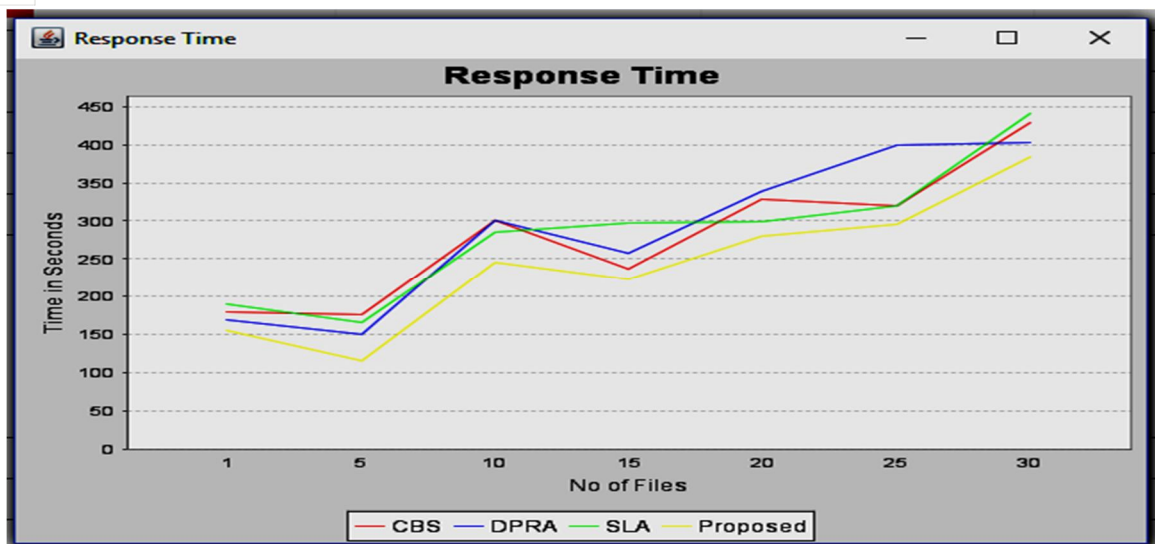
Figure:- 'Response Time' Graph

VII.   CONCLUSION

The evaluating parameters are used to check the performance of the framework with the existing systems. The proposed work shows good results than existing systems. In terms of throughput, the proposed work shows little higher values than existing systems. But for other parameters the proposed work shows better results than the existing systems. For evaluating parameters such as CPU Utilization, Energy Efficiency, Scheduling Time, Response Time the proposed work is far more better than existing systems. The table from 2 to 6 shows the values generated by the framework which indicated the better performance of the proposed work. The respective graphs from 1 to 5 shows the comparative performance of the proposed work and the values suggest that the 'Yellow line' which represents the proposed work has better values or results than the existing systems.

REFERENCES

[1]   B. Hayes, "Cloud computing," Communication ACM, vol. 51, no. 7, p. 9, Jul. 2008.

[2]   R. G. Tiwari, S. B. Navathe, and G. J. Kulkarni, "Towards Transactional Data Management over the Cloud," 2010 Second Int. Symp. Data, Privacy, E-Commerce, pp. 100–106, Sep. 2010.

[3]   Qi Zhang, Mohamed FatenZhani, RaoufBoutaba, Joseph L. Hellerstein, "Dynamic Heterogeneity-Aware Resource Provisioning in the Cloud", IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2014

[4]   BalajiPalanisamy, Aameek Singh, Ling Liu, "Cost-effective Resource Provisioning for MapReduce in a Cloud", IEEE Transactions on Parallel and Distributed Systems, 2014.

[5]   L. Minas and B. Ellison, Energy Efficiency for Information Technology: How to Reduce Power Consumption in Servers and Data Centers. Intel Press, 2009.

AUTHOR PROFILE

Mr. Dheeraj Bhimrao Lokhande.

He received his BE degree in 2008 from walchand Institute Technology from Solapur University,Solapurwith first class with Distinction. He Qualified GATE Examination in 2008. He Completed his ME(CSE) postgraduate degree in 2012 from Solapur University, Solapur with First Class. In 2014,He Registered for Ph.D(CSE) from Solapur University,Solapur, Maharshtra, INDIA.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089   (24*7 Support on Whatsapp)