Global Journal of Computer Science and Technology: B
Cloud and Distributed
Volume 14 Issue 2 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Enhancement of Map Function Image Processing System using DHRF Algorithm on Big Data in the Private Cloud Tool

By Mr. U. Mehraj Ali. & Dr. A. John Sanjeev Kumar
Thiagarajar College of Engineering, Madurai, Tamilnadu, India

Abstract- Cloud computing is the concept of distributing a work and also processing the same work over the internet. Cloud computing is called as service on demand. It is always available on the internet in Pay and Use mode. Processing of the Big Data takes more time to compute MRI and DICOM data. The processing of hard tasks like this can be solved by using the concept of MapReduce. MapReduce function is a concept of Map and Reduce functions. Map is the process of splitting or dividing data. Reduce function is the process of integrating the output of the Map's input to produce the result. The Map function does two various image processing techniques to process the input data. Java Advanced Imaging (JAI) is introduced in the map function in this proposed work. The processed intermediate data of the Map function is sent to the Reduce function for the further process. The Dynamic Handover Reduce Function (DHRF) algorithm is introduced in the reduce function in this work. This algorithm is implemented in the Reduce function to reduce the waiting time while processing the intermediate data. The DHRF algorithm gives the final output by processing the Reduce function. The enhanced MapReduce concept and proposed optimized algorithm is made to work on Euca2ool (a Cloud tool) to produce an effective and better output when compared with the previous work in the field of Cloud Computing and Big Data.

Keywords: cloud computing, big data, mapreduce, euca2ool, DHRF algorithm.

GJCST-B Classification : D.4.2, I.4.10

Strictly as per the compliance and regulations of:

Enhancement of Map Function Image Processing System using DHRF Algorithm on Big Data in the Private Cloud Tool

Mr. U. Mehraj Ali. [α] & Dr. A. John Sanjeev Kumar [σ]

Abstract- Cloud computing is the concept of distributing a work and also processing the same work over the internet. Cloud computing is called as service on demand. It is always available on the internet in Pay and Use mode. Processing of the Big Data takes more time to compute MRI and DICOM data. The processing of hard tasks like this can be solved by using the concept of MapReduce. MapReduce function is a concept of Map and Reduce functions. Map is the process of splitting or dividing data. Reduce function is the process of integrating the output of the Map's input to produce the result. The Map function does two various image processing techniques to process the input data. Java Advanced Imaging (JAI) is introduced in the map function in this proposed work. The processed intermediate data of the Map function is sent to the Reduce function for the further process. The Dynamic Handover Reduce Function (DHRF) algorithm is introduced in the reduce function in this work. This algorithm is implemented in the Reduce function to reduce the waiting time while processing the intermediate data. The DHRF algorithm gives the final output by processing the Reduce function. The enhanced MapReduce concept and proposed optimized algorithm is made to work on Euca2ool (a Cloud tool) to produce an effective and better output when compared with the previous work in the field of Cloud Computing and Big Data.

Keywords: cloud computing, big data, map reduce, euca2ool, DHRF algorithm.

I. Introduction

Cloud Computing is a well known concept also known as, IT on Demand. Cloud computing is the concept, which is run by most of the IT companies like e.g. Amazon's EC2, Intel's Intel Manager. Cloud computing package has a set of clients and servers. The packages are powered with high storage capacity, high flexibility and high computing performances. Cloud computing has the several concepts in it like, cloud storage and cloud security; cloud storage reaches to the users a lot. Euca2ool is a Cloud tool, proposed in this work is to enhance the uses of this tool; where the user has to create an account in the Eucalyptus Partner Cloud (EPC). Since this tool can easily interact with Amazon's EC2 on any operating system, this Ecu2ool has been selected in this work.

Java Advanced Imaging (JAI) is regularly used for encoding and decoding the image that is saved in the HDFS. JAI is introduced mainly for the betterment of imaging processing process. The job of the JAI is to translate or to transform the image format. The concepts of Big Data, JAI, Hadoop and Map Reduce functions are proposed in this work. This work is executed with the private cum hybrid cloud tool called Euca2ool.

The Fig. 1 shows the full Architecture of the proposed work. Two set of processing techniques are applied over the input and Fast corner_9 corner detection is also used in the system. Presently, this regular set of work is made with the other corner detection method and scheduling algorithm for 2D to 3D data processing [1]. In the proposed work, it has been proved that, there is another better corner method, improved SAD matching and Scheduling DHRF Algorithm, which could benefit the client in the useful manner. DHRF algorithm has proved that, it works better than the existing algorithm while processing the Reduce function. One of the main concepts in the proposed work is introduction of JAI used for image transformation. If the images are in the improper format, JAI changes the image to the right format.

Author α: Department of Computer Application, Thiagarajar College of Engineering, Madurai, Tamilnadu, India.
e-mail: mehrajtce@gmail.com

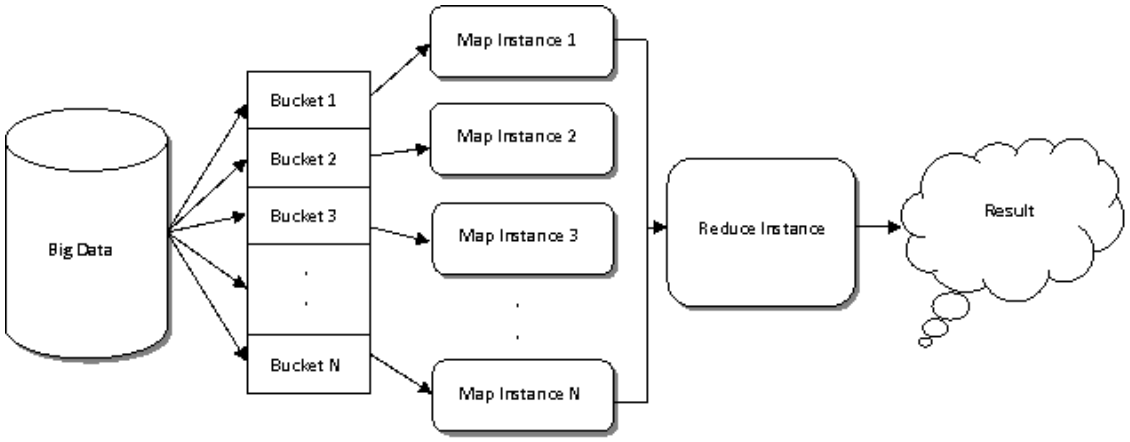

Figure 1 : Architecture of Hadoop for Map Reduce Function

The Fig. 1 shows the full Architecture of the proposed work. Two set of processing techniques are applied over the input and Fast corner_9 corner detection is also used in the system. Presently, this regular set of work is made with the other corner detection method and scheduling algorithm for 2D to 3D data processing [2]. In the proposed work, it has been proved that, there is another better corner method, improved SAD matching and Scheduling DHRF Algorithm, which could benefit the client in the useful manner. DHRF algorithm has proved that, it works better than the existing algorithm while processing the Reduce function. One of the main concepts in the proposed work is introduction of JAI used for image transformation. If the images are in the improper format, JAI changes the image to the right format.

The raw data formats input are converted to the fixed frame size and then the data compression is done. The compressed data is scaled to a fixed frame size. The received output will be a better one with high flexibility, less time estimation and less error percentage. Mostly the medical data will be in the DICOM format and rarely in the JPEG format and the output received will be in the .jpeg format.

According to the literature, the best way to process the large volume of data is enhancing Hadoop. Hadoop basically has HDFS (Hadoop Distributed File System) and Map Reduce function. The enhancement of Hadoop over Cloud Computing will obtain additional improvement in result and also in the waiting time. Hadoop can solve the problems like failure of Machines, fault tolerance, which occur during the processing of large set of data. These minor problems may result in the failure of the work. So, in order to overcome these problems, Hadoop is used.

When the speed of the machine did not match with the Map Reduce function, the tasks will be delayed [3]. To rectify this problem, the Ecu2ool is used to manage the works of the Map Reduce function. To reduce the waiting time during the operation, a Dynamic Handover Reduce function (DHRF) algorithm is implemented in the Reduce function. The proposed algorithm will work on the reduce function to reduce the waiting time during the operation.

If there are a few small files, then Map task processes only a little amount of input data, and as a result, there are many unscheduled Map tasks [4]. The concept of unscheduled Map task is common. In the proposed work, the Map task are unscheduled, the process of intermediate data to the Reduce function will be a little hard task. So, according Tin-Yu Wu et al [1], a switch algorithm was proposed. But, since there may occur some technical fault while processing the Reduce function, Handover technique is implemented. So that, the process will wait, until each process gets over completely. The HDFS system is utilized in this process. The cases may be either the small pieces of data or the large volume of data; the HDFS storing system supports all the formats for storing the data.

This paper reports the experimental work on big data problem and its optimal solution using Hadoop cluster, Hadoop Distributed File System (HDFS) for storage and using parallel processing to process large data sets using Map Reduce programming framework [5]. Big data chunks with different size and sequence will be computed in each node, so that transfer of a chunk is overlapped with the computation of the previous chunk in the node, as much as possible [6]. The data transfer delay can be comparable or even higher than the time required for computing the data [7]. Use of Dynamic Switch of Reduce Function (DSRF) algorithm, a scheduling scheme on the Reduce functions for users who compute simultaneously to acquire Reduce resources to finish the tasks efficiently [8].

In the proposed work, Dynamic Handover Reduce function (DHRF) algorithm is implemented in the Reduce function where the system waits until the process get over. The proposed module converts the size and format of the image using the following options:

max Width, max Height, Image Format [2]. JAI is introduced with small change in the coding. If max Width, max Height, is the case, it would take more space in HDFS. This may result in occupation of more space. The Reduce function will wait for the Map function to generate intermediate data; the Reduce function can switch to another task to combine the image data first [1]. The case of incompletion may occur while processing the Reduce function in the existing work, when the machine get faults like virus attacks, corruption of Operating System and others.

a) Hadoop Distributed File System (HDFS)

Hadoop Distributed File System is the file storing section of Hadoop. Fig. 2 shows the format, by which the inputs in the various forms or formats are stored. Those inputs or the data gets stored in HDFS to process the Map Reduce Function. HDFS handles both the structured and unstructured data. The Hadoop's job is to split the data and distribute to the hosts to compute. This work will be done simultaneously in the parallel manner. This is called as distributed and parallel computing. The host depends upon the strength and the storage of the computing capacity.

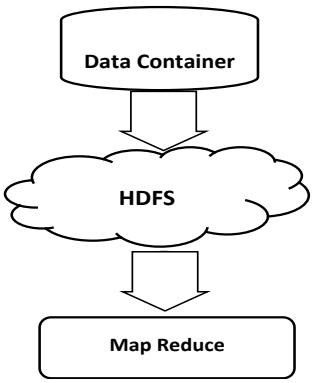

Figure 2 : Modules in MapReduce Function

II. Data Processing Techniques

In the proposed work, two various types of image processing techniques are applied for the input (Big Data) as shown in the Fig.3. The first technique is the combination Grayscale and Sobel edge detection. The second technique is the combination of Gaussian Blur and Fast Corner detection method. The Grayscale conversion is the best method to convert the original image into Black and White image and Sobel edge detection is the method of detecting the edges of the images.

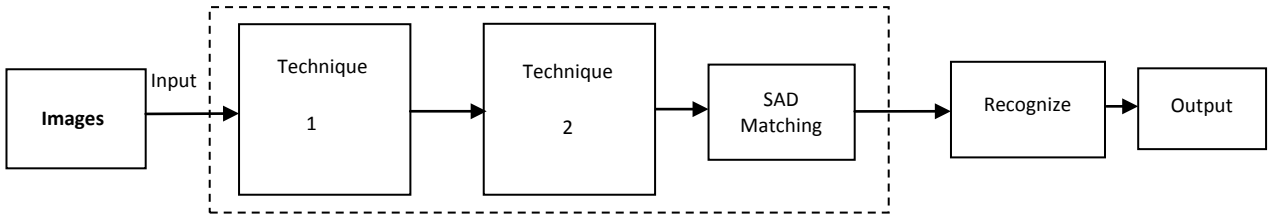

Figure 3a : Sketch of Data Processing Techniques

Gaussian Blur is the best blurring technique ever, used here to blur the input image. Finally the Fast Corner Detection technique, since that is the best technique. After the completion of both the processing, the data is then transferred to the next level of Matching called SAD matching. Then the integration of the intermediate data is done to recognize the output.

a) Harris corner detection

The Harris corner detection is used in case when there is a correlation in the threshold of data. A detection method which could give a better result than the Harris corner is, Fast corner_9 which is shown in Fig. 4. Fast corner_9 has been applied in the proposed work, to fine tune the result than the Harris corner, Susan, zheng and harr [9] methods.

$$E(u,v)=\sum_{x,y} w(x,y)\left[I(x+u,y+v)-I(x,y)\right]^2$$

The given equation finds the solution for the Harris corner method with the argument (u,v).

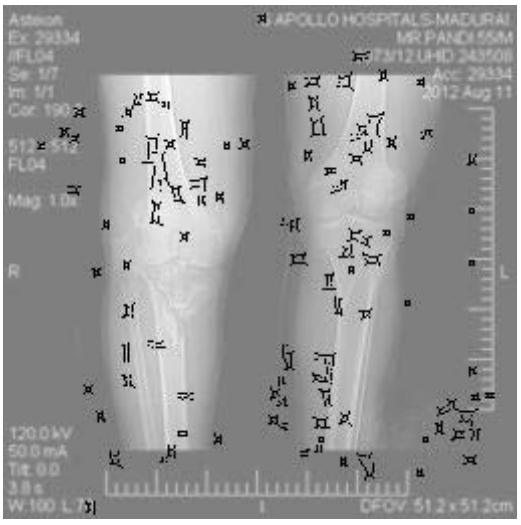

Figure 3b : Threshold data obtained through Harris Corner Method

In Harris corner, a square mask with point *p (x, y)* was established. When the mask of Grayscale value is higher than the threshold, the point P is defined as the corner. Here *E* denotes the patch area, *I* denotes the given image. Where *(u,v)* denotes the image patch point and while processing, *(x,y)* get shifted from *(u,v).*

b) Fast Corner 9

In the presence of various Corner detection methods include Harris corner method, Susan, zheng and harr. Among these methods, Fast corner_9 is used due to higher clarity when compared with other corner detection method. This Fast corner detection has come from Moravec (SSD) and the Harris Corner Detection (second derivative of SSD). By using the non minimal suppression it is determined. This is calculated by subtracting the original value, by comparing the *p* value and with *t* value. This has been implemented on Client Machine with the configuration Intel (R) Core 2 duo, 4 GB RAM and 2.93 GHz processor and the Server with the Configuration of IBM X 3400 M3 Server, Intel Xeon E 5507 (Quad Core), 2.26 Ghz Processor, 146 GB DDR3-1333 Mhz ECC RAM, 146 GB 10K SAS Hard Disk Drive.

$$v = min\begin{cases}\sum(p - pixel\ value)n + if\ \ \frac{1}{n}\ (value - p) < t\\ \sum(pixel\ value - p)n + if\ \ \frac{1}{n}\ (p - value\ ) > t\end{cases}$$

Where *p* denotes the Grayscale value and *t* denotes the Threshold value.

c) Sum of Absolute difference (SAD)

SAD is a mathematical term that has been enhanced here to find between differences the two P blocks of the processed data.

$$SAD = \sum_{(u,v)\epsilon\, w}[A_1 1(u,v) - A_2(x+u, y+v)]$$

Where, A_1 and A_2 are the two images. Like the Harris Corner method, here also (u,v) denotes the image patch area and while processing, (x,y) gets shifted from (u,v). Before the implementation of SAD there is an existence of SSD (Sum of Squared Differences) to find the difference. As SSD is an old, it isn't able to produce a clear result like SAD. The Relationship between the intersection points detected by Harris corner detection method and labels the corresponding points for computers to judge the corresponding locations of intersection. In relationship between the corners, the interpolations are used to judge the location of the corner while being photographed to simulate the data.

III. Proposed Work

Euca2ool is a Cloud tool that is private cum hybrid Cloud Tool. Though there are many tools to solve the problems in big data.. This enhanced tool will definitely give the better result in both the time and cost estimation. Basically Hadoop is enhanced of HDFS and Map Reduce function. Since the Map Reduce function is an open source, the codes can be edited and modified. The coding is edited now in the Map part. The coding is made in such a way that, the inputs are split into maximum of ten parts. So that the Map function will be easy while processing. This change in the Map function will reduce the processing time.

The coding or the application set up of the image processing techniques is installed on the Euca2ool, to run the experiment. Basically Map is the primary function in the process of Big Data. Splitting the data into maximum number of part is called as Map Function. In the proposed work, Pre-Map concept is implemented. Pre-Map is the concept of Map the Map function. The situation in which, when the Map function is proposed to do, the data will be mapped before the Map function into maximum number of parts.

a) How Cloud Computing is involved in Image Processing in the proposed work?

The HDFS is used to store, retrieve and process the data. Few image processing techniques are used in the cloud computing. Since, Big Data is referred to the context called image/data. When the data is taken as the input, it is made to undergo the Map Function. The function of the Map function is to Split the data into maximum number of data called as intermediate data. Those intermediate data are split to process further. Two sets of image processing techniques are used in the Cloud Computing.

JAI does the task of converting the data. Either the data may be a structured or unstructured one. By using the JAI, the data are converted to structured data. Fig 4 shows the results of obtained from the data processing techniques. Further, the inputs are processed first with the Pre-Map technique. Pre-Map technique is used, to split the data before sending the data to Map function. This Pre-Map function split data in to maximum number of parts, in order to reduce the processing time during the Reduce application.

After the function of Pre-Map and Map Function gets over on the Data, the data are set to intermediate data. The Map function is done by the client machines. These intermediate are made to undergo the two set of image processing techniques. To avoid the waiting time, the Dynamic Handover Reduce Function (DHRF) is applied on each and every intermediate data. This results in the output of the input (Big Data).

Figure 4 : Results of Data obtained from image Processing Techniques

IV. DHRF Algorithm

DHRF (Dynamic Handover of Reduce Function) is an algorithm implemented in the proposed work. The function of this algorithm is to reduce the waiting time during the Reduce function. The work of the reduce function is to integrate the processed data. Two sets of image processing techniques are applied on the intermediate data such as Grayscale, Sobel, Gaussian and Harris Corner detection. These techniques are applied on the splitted data which are produced after the application of Map function. Each and every intermediate data are applied with two techniques.

```
BufferedImage res = new BufferedImage(width,height,BufferedImage.TYPE_BYTE_GRAY);
                    // Initialize the image process
byte[] bytesCompressed = compressor.compress(imagetoCompress);
Deflater deflater = new Deflater();
deflater.setInput(bytesToCompress);
                    // Produce the data compression
BufferedImage resizedImage = new BufferedImage(IMG_WIDTH, IMG_HEIGHT, type);
Graphics2D g = resizedImage.createGraphics();
g.drawImage(originalImage, 0, 0, IMG_WIDTH, IMG_HEIGHT, null);
g.dispose();
                    // put the data into scaling
static{
   URL.setURLStreamHandlerFactory(new FsUrlStreamHandlerFactory());
}
                    // write the map reduce structure
 in = new URL(PATHTOBEMAPPED).openStream();
IOUtils.copyBytes(in, System.out, 2, false);
                    // set the server to handle mapper
FSDataOutputStreamout=fileSystem.Create(path);
   InputStream in = new BufferedInputStream(new new File(source)));
                    // mark data into HDFS of Hadoop
Process the image until completing the grayscale,sobel, guassian, fast corner, SAD matching of the image.
                    // operate the data process until the data processed
```

map(in_key, in_val) -> list(out_key, intermediate_val)

reduce(out_key, list(intermediate_val)) -> list(out_value)

// Set the MapReduce Operation

FileSystem fs = file.getFileSystem(context.getConfiguration());

FSDataOutputStream fileOut = fs.create(new Path("your_hdfs_filename"));

// write the data mapper

reduce(WritableComparable, Iterator, OutputCollector, Reporter)

continue until reducer task is complete

// send mapper output data to reducer

JobConf.setNumReduceTasks(int)

// set small unit value to the task and reducer wait queue

interrupt.task

store the result(image).

When the image is processed by the template, the image is compressed and scaled then produces the Map functions. In the first set of Grayscale method, it delivers the image in black and white without noise disturbances. The Grayscale image output handovers the edge detection technique to the next process. It removes the outer layer of noise disturbance. So enhancement of the best edge detection technique called Sobel edge detection technique is done.

Then the implementation of Gaussian blur reduces the image noise. It is for pre-processing stage for any image enhances structure. So the retrieval of formal blur image can be used for the detection of corners. The corner method defined as a point of two different edge directions and dominant. In this, corner detector works uses only a segment test, so the result is very accurate and quality at the mean time of time reducing in the .jpeg format.

Finally Map function results are sent to reduce operation. The implementation of DHRF algorithm focuses on Reduce function integrating the task and allots the process to produce the result .Jpeg format. In Reduce function, it involves the small unit of value for determining the task and sequentially it completes the process. Whether the process is completed, it produces the result .jpeg format.

V. Result and Discussion

Fig 5 shows the result of the Fast Corner method. In this figure, the sharp edges and the damaged parts are pointed out sharply. The red mark shows the infected corners. When compared with the Harris corner method in the existing work [2], this is proved to be the best corner detection. The application of the two image processing techniques, gives the expected result.

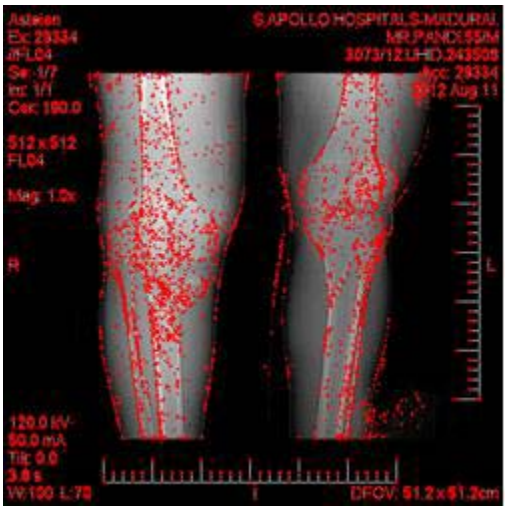

Figure 5 : The result shows the sharp edge and damaged part through Fast Corner method

The main advantage during the execution was, the output of the data is marked with better quality and there is the usage of Fast corner_9 method to detect the corners of the image. While comparing with existing, proposed Fast corner_9 method shows the corner more clearly. During the comparison, it has been proved that, the method applied in the proposed algorithm works better than the existing. Whatever the size or format of images may be, the result will be produced in the pre defined format. The result occupies less space when compared with the size of the input for storing.

The Fig 6 graphical representation is between the estimated time and total number of Machines. In the proposed system, the taken to process the image is comparatively less than that of existing system.

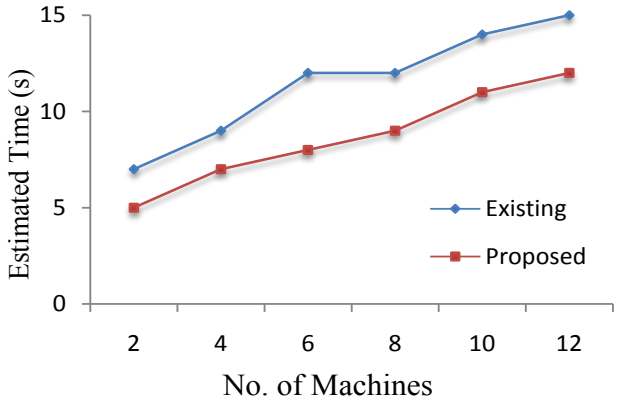

Figure 6 : Effect of proposed and existing techniques with respect to time and No. of machines

Finally, this Fig 7 shows the graphical representation between Existing Map function with the Proposed Map function. In the existing Map function, the data will simply just get splitted in the normal way. But in the proposed system, we have introduced Pre-Map technique, it basically Maps the data before sending to the Map Function. By default the data processing will be simple. This is graphically explained in the bellow picture.

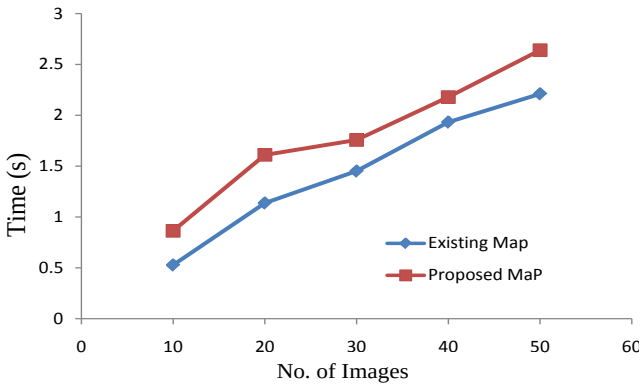

Figure 7 : Effect of mapping techniques with respect to time and No. of image

VI. Conclusion

Already an existing work was done with four data processing techniques and with an algorithm was implemented in the Reduce function. But in this proposed system, the processing techniques have reduced from four to two image processing techniques, with proposed an optimized scheduling algorithm. This work resulted with less waiting time and error percentage. An application of JAI and Pre-Map technique with Hadoop over Ecua2ool results with far better result in the entire manner when compared with the existing system. Further, our research works will be on application of scheduling over the Heterogeneous Networks for scheduling and resource allocation.

References Références Referencias

1. Tin-Yu Wu Chi-Yuan Chen, Ling-Shang Kuo, Wei-Tsong Lee, Han-Chieh Chao, " Cloud-Based Image Processing System With Priority-Based Data Distribution Mechanism", *Computer communi -cations*, COMCOM 4676, 2012.
2. Hyeokju Lee, Myoungjin Kim, Joon Her, and Hanku Lee, "Implementation of Map Reduce-based Image Conversion Module in Cloud Computing Environment", 978-1-4673-0250-0/12/$31.00 ©2012 IEEE.
3. Patel, A.B. ; Birla, M. ; Nair, U., "Addressing big data problem using Hadoop and Map Reduce, Engineering (NUiCONE)", Nirma University International Conference on Digital Object Identifier, 10.1109/NUICONE.2012.6493198 Page(s), 1 – 5, 2012 .
4. Gueyoung Jung ; Gnanasambandam, N. ; Mukherjee, T., "Synchronous Parallel Processing of Big-Data Analytics Services to Optimize Performance in Federated Clouds, Cloud Computing (CLOUD)", IEEE 5th International Conference on Digital Object Identifier, 10.1109/CLOUD.2012.108, , Page(s), 811 – 818 Publication Year, 2012.
5. HaoYu, "FAST corner detection-Machine learning for high speed corner detection", Publication Year 2010.
6. Shini.S.G , Dr.Tony Thomas ,Chithraranjan.K, "Cloud Based Medical Image Exchange-Security Challenges", Procedia Engineering 38 (2012) 3454 – 3461, Publication Year, 2012.
7. Nauman Sheikh, "Chapter 11 – Big Data, Hadoop, and Cloud Computing, Implementing Analytics A Blueprint for Design, Development, and Adoption", Pages 185–197, Publication Year 2013.
8. Mark D. Ryan, "Cloud computing security, The scientific challenge, and a survey of solutions", The Journal of Systems and Software xxx 2013.
9. Jules J. Berman, Chapter 8 – "Simple but Powerful Big Data Techniques, Principles of Big Data Preparing, Sharing, and Analyzing Complex Information" Page(s), 99–127, Publication Year 2013.

This page is intentionally left blank