International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Big Data Analytics for Images in Public Cloud using Map Reduce on Local Clusters

Buvaneswari.V.B, S.Shanthi, M.Pyingkodi

Abstract— MapReduce is a programming model used for parallel computing of big data in public cloud. Big Data have characteristics like variety, velocity and volume. The research work carries out MapReduce using Matlab which is a powerful image processing and numeric computation tool. The research considers unstructured image data in public cloud Dropbox as big data and applies MapReduce algorithm to map and reduce all the images stored in it. The research work aims to retrieve the images in public cloud with maximum Red, Green, Blue color and the colors that intersect between them. The same code is modified to find all Red, Green and Blue that supports more parallelism and aids in improving the speed of MapReduce by eliminating the dependency between iterations. The speed of parallel MapReduce shows considerable improvement only with increased file size and coding style. Parallel MapReduce computation is carried out with default workers, three and four workers of the local cluster with scale up architecture. This model is developed using Matlab and can be implemented in Hadoop as well.

Keywords: MapReduce, Big Data, Parallel Computing, Cloud, image processing, cluster

I. INTRODUCTION

LinkedIn etc makes available huge volumes of images that grows rapidly. Multi-platform mobile messaging applications like Whatsapp messenger [9] and Hike Messenger [4] make volumes of images available easily with the help of smartphones and personal computers. Internet, Various Sensors and digital processing also create tons of image data. A technology that is powerful and cost effective is needed to handle oceans of data. Image storage clouds such as Flickr[2][21], Picasa[10] ,Google Photos [5] and Dropbox[1] helps in storage of image data automatically through auto backup from smartphones and personal computers. Flickr allows one TB free space for image storage with maximum file size of 200MB. Google Photos talks about unlimited free storage. Dropbox free storage ranges from 15 GB to 50 GB.

Manuscript published on November 30, 2019.
* Correspondence Author
Buvaneswari.V.B*, Assistant Professor P.G and Research Department of ComputerScience,Government Arts College, Coimbatore, Tamil Nadu, India.
Dr. S.Shanthi, Department of Computer Applications, Kongu Engineering College, Erode, India. (Email: shanthi.kongumca@gmail.com)
M.Pyingkodi Department of Computer Applications, Kongu Engineering College, Erode, India. (Email: pyingkodikongu@gmail.com.)

1.1 Big Data Characteristics

Big Data[7] demands Information Processing that is cost and innovative that helps in making decision and automating process[3]. The characteristics of Big Data are

- Volume – Large data. It can be terabyte, petabyte , exabyte, zettabyte or yottabyte.
- Velocity – The speed or rate at which data enters handled at particular time.
- Variety – Variety deals with different type of data that Big data supports. In includes text, image, audio,video, social media data.Log file, sensor data etc.It supports both structured and unstructured data[22].
- Variability - peak data loads.
- Complexity- Linking data from different sources followed by cleansing and transformatio

1.2 Advantages of Big Data

Big Data facilitates faster and better Decision Making. It performs Market Analysis and provides new products and services based on it. Big Data technologies helps in acheiving substantial Cost reduction and Time saving through improved efficiency and supports Automated Processing.

1.3 Challenges and Remedies for Big Data

- Fitting into Memory – Solved by allotting data in chunks to memory
- Long processing time – Can be tackled using parallelism
- Fast streaming brings difficulty in storing data – Handled through cloud storage
- Security and Privacy concerns – Dealt using encryption and decryption algorithms

1.4 MATLAB

Matrix Laboratory (Matlab) is a language with high performance suitable for numerical computing, parallel computing with multiprocessor and multicore, programming and visualization [23]. Its highly suitable for image processing

1.4.1Advantages of using Matlab for Big Data Analytics over Hadoop

- Easy installation
- Lower

Hardware requirements

- User friendly
- Easily handles image datastore.
- Doesn't need exhaustive knowledge to operate
- Sophisticated data analysis
- Minimize communication and computation through cores
- Automatic load balance
- Supports both scale up and scale out Architecture [26]

1.5 Big Data and Cloud

Big Data and Cloud are made for each other. Terabytes of data are offered free.Dropbox is adopted in this

Fig.1 DropBox Cloud

Fig.1 shows Dropbox cloud with different files, shared folders and team folders.

research work because Dropbox folders can be shared and images does not lose its rich detail as in other clouds due to compression and decompression. MapReduce is a significant programming model that is suitable for cloud computing [15]. Cloud offers an open environment where knowledge can be easily shared[31].Variety of cloud architectures are provided by various cloud providers [24]. Scalable architecture is provided by certain providers [30]. The storage service for Big Data Analytics is given by cloud vendors[14]

II. RELATED WORKS

Mapreduce is suitable model for shared memory systems with performance that is scalable and code that is simple [27]. Phoenix implements map reduce and manages creation of threads, task scheduling in a dynamic manner, partitioning data and fault tolerance at runtime in processor nodes [27]. Its performance and error recovery capabilities are assessed [27].

The MapReduce doesn't need schemas like DBMS and suitable for processing unstructured data and can be described in terms of key value pairs [25].

The features such as Elasticity, pay-per-use, low upfront investment, low time to market and transfer of risks encourage the usage of economically feasible cloud computing [13]. Taking this into account the research work uses Dropbox cloud.

Security and privacy are important issues in Big Data because third party services and infrastructures are used for processing [21] [16]. Simple functionalities are only provided by key-value stores [13].

Internal communication of multicore is less compared to external multiprocessor communication. So the research work considers internal communication with dual core**.** Speedup is calculated with 2, 4, 8 and 16 processing cores for the same algorithm. Speedup increases with number of cores**.** Due to insufficient cores the research work considers differentworkers for the same algorithm. M and R should be more than number of workers for dynamic load balancing [19]. In case if one worker fails the work will be distributed to other workers [18]. Static scheduling will lead to resource utilization that is unoptimized. Each reducer should be allotted data in equal proportion otherwise one reducer will be running more time and others will sit idle

III.PROBLEM DEFINITION

The aim is to find the image with Maximum Red, Green, Blue, Yellow, Magenta and Cyan among all the images that are available in public cloud DropBox.

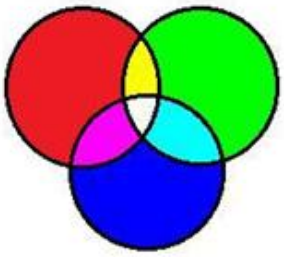

Fig 2. RGB color components

The above fig.2 shows Red, Green and Blue Color along with intersecting colors Magenta, Cyan and Yellow.

Select Red Pixel, Blue pixels and Green pixels . For Red pixel components of Red is high whereas components of Blue and Green are low. Similarly for Green pixel, Green component is high and both the components of Red and Blue are low. The same holds for Blue pixel too in which components of Blue are high and components of both Red and Green are low. RGB (255, 0, 0) is not sufficient to find the red color and its shades. Likewise RGB (0, 255, 0) is not enough to identify the Green color and its shades. RGB (0, 0,255) can identify Blue color but not the shades of blue.

To identify Blue color and its different shades in an image apply the formula

Blue = B – Maximum(R,G) (Eqn 1) and applying it for different shades of Blue give

RGB (32, 178, 170) = -8
RGB (0, 139, 139) = 0
RGB (0, 255, 255) = 0
RGB (127, 255, 212) = -43
RGB (0, 191, 255) =64
RGB (135, 206, 235) = 29
RGB (135, 206, 250) = 42
RGB (0, 0, 139) = 139
RGB (0, 0, 255) = 255
RGB (65, 105, 225) = 120
RGB (0, 0, 128) = 128
RGB (106, 90, 205) = 99

RGB (147, 112, 219) =72
RGB (148, 0, 211) = 63
RGB (186, 85, 211) = 25
RGB (128, 0, 128) = 0
RGB (218, 112, 214) = -4

By applying Eqn.1 the image with maximum blue pixel values is selected from Public cloud Dropbox is Blue = 226.8745- values is selected from Public cloud Dropbox is Blue = 226.8745- Similarly, to obtain maximum Red and Green use Equation 2 and Equation 3.

Red = R– Maximum(B,G) (Eqn 2)

Maximum Red image from Dropbox cloud is Red =254 – Maximum(0,0)=254 Green = G – Maximum(R,B) (Eqn 3)

Maximum Green image retrieved from Dropbox cloud is Green =255-Maximum(0,1)=254

The intersection of color components Magenta, Yellow and Cyan are obtained by using Equations 4, 5 and 6.

Red Blue (Magenta) = R + B –G (Eqn 4)

The maximum Magenta image values for the Dropbox cloud image is Magenta=238.1480+145.2114-98.4214=284.9380

Red Green (Yellow) =R + G –B (Eqn 5)

The maximum Yellow image values for the Dropbox cloud image is Yellow=253. 7440+224. 2546-1.9993==475.9993

Green (Cyan)=B+G – R (Eon 6)

The maximum Cyan image values for the Dropbox cloud image is Cyan=252.6810+2535918+-138.8706=367.4022

IV IMPLEMENTING MAPREDUCE USING MATLAB

4.1 Datastore

Data collections can be accessed in a chunk based manner that fits in memory using data store. Data store read and analyze data in chunks. It acts as a repository for large data collections that cant be stored in memory at a stretch. Datastores can be created for text, image, keyvalues etc.

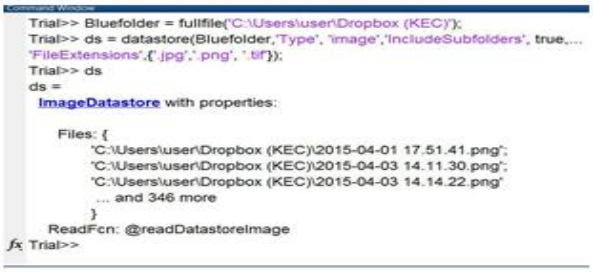

Fig.3 Reading from ImageDataStore

The fig.2 shows tha timages are read from ImageDatastore that consist of 349 images. Here Dropbox is considered as ImageDatastore . The images are extracted from public cloud Dropbox with fileextensions jpg, png and tif.

4.2 MapReduce

Fig.4 Working of Map Reduce in Dual Core with Default Workers

Fig.4 depicts the input image image is given to the mapper through worker 1 and worker2 . The mapper produces intermediate data which is given to the reducer which produces the desired output.

4.2.1Algorithm for MapReduce

1. MapReduce reads image data in chunks from data store. In our example from Dropbox
2. .Key-value pairs are added to KeyValueStore which is a data storage object that is intermediate.
3. Based on unique key MapReduce groups all values in KeyValueStore.
4. For each unique key reduce function is called. Unique key and its associated values are passed to ValueIterator Object.
5. Aggregate intermediate results after iterating through the values from map function.
6. Reduce function adds final key-value pairs.

4.2.2 MapReduce Progression

The MapReduce progression maps the tasks to map function which is reduced by the reduces function. The number of map tasks is always greater than the reduce tasks.

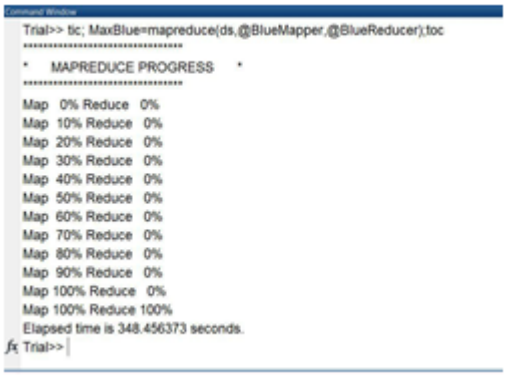

Fig. 5 MapReduce Progress

Fig.5 depicts the mapreduce progress. It has two functions BlueMapper and BlueReducer that acts as Mapper and Reducer respectively.

4.2.3 Key Value DataStore

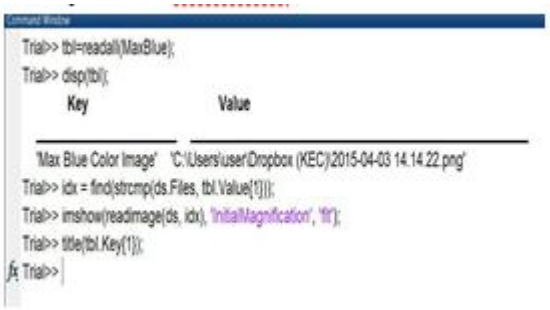

Fig. 6 Final Key-Value Pair

The fig.6 shows the key 'max Blue Color Image' and the value of the location of the image which has maximum blue color value.

4.2.4 Image Corresponding to Key-Value Pair

Fig 7. Maximum BlueColor Image obtained from Cloud

Fig. 7 shows the maximum blue color image that is extracted from Dropbox cloud from 349 images. The key values for this image are given in fig. 6.

4.2.5 Parallel Computing

Parallel computing can be performed using parallel pool with 2 workers. In the Dual Core processor the two cores act as 2 workers as default. But more workers can also be added with little improvement in performance. Multi core improve performance. For eg. If Quad core processor is used 4 default workers can run in parallel. If quad core with hyperthread is supported by machine(i7 processors support hyperthreading) then 8 default n workers can run using quad core processor. Multi cores are preferred over multi processes because communication cost is low in multi core compared to multiprocessor. Internal communication cost is low in multicore compared to external cost.

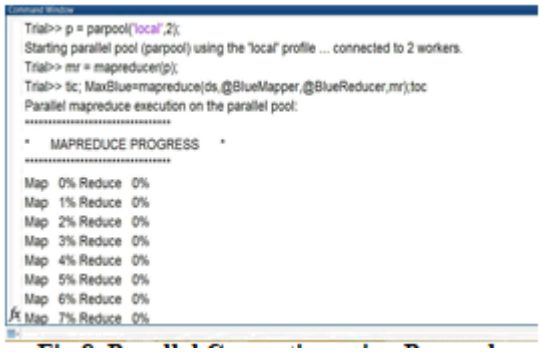

Fig 8. Parallel Computing using Parpool

Fig.8 depicts that parallelmapreduce is carried out using parpool which is a parallel pool. Here local profile is used with 2 workers connected to it. The Blue color square at the bottom depicts that parallel computing is in process.

4.2.6 Computing with 64 bit

The 64 bit Matlab version facilitates fast computing compared to 32 bit Matlab. 64 bit matlab needs 64 bit OS which gives 4000 times more address space compared to 32 bit OS.

4.2.7 LOCAL CLUSTER

The Matlab Client session is run by local scheduler. When a job is submitted to local cluster [8] for evaluation the scheduler allots tasks for each **worker** permitted in the local profile. If there are more tasks that cannot be allotted to the available workers, the scheduler waits for the tasks to complete before permitting the workers to take up new task. The number of workeres in the local profile can be altered. The default workers will be equal to the number of cores available on the machine. But maximum of 512 workers can be supported using parcluster with parpool.

Fig 9. Local Cluster and associated jobs

Fig.9 Shows the properties and jobs associated with local cluster. The default number of workers available is 2 because the computer has only dual cores. The number of workers can be extented too.

V. PERFORMANCE WITH KEY-VALUE

PAIRS & RESULTS

The Average Elapse time is the average of 10 runs

5.1 Exp #1: Single Key Value - finding maximum Blue color

Fig.2 show there in one reduce operation and this does not improve the performance. Moreover finding maximum blue color does not support parallellism much. The value of the consecutive iterations are dependent on previous iterations.

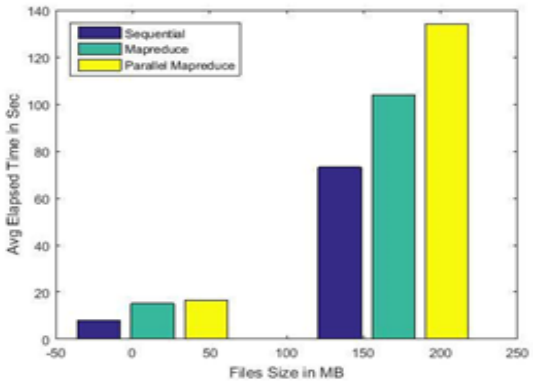

Fig.10 Finding Maximum Blue Color using Single Key-value

Fig.10 Depicts the Elapsed time of Sequential, MapReduce and Parallel MapReduce. There is vast difference in performance between sequential, MapReduce and Parallel MapReduce

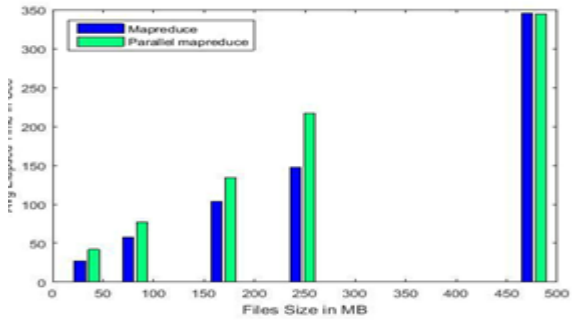

Fig.11 Finding Maximum Blue Color with Single Key-value using Increased File Size

Fig.11 Shows that even improvement in file size does not bring out improvement in performance .

5.2 Exp #2: Three Key Value - finding maximum Red, Green and Blue.

```
Map  0% Reduce  0%
Map 10% Reduce  0%
Map 20% Reduce  0%
Map 30% Reduce  0%
Map 40% Reduce  0%
Map 50% Reduce  0%
Map 60% Reduce  0%
Map 70% Reduce  0%
Map 80% Reduce  0%
Map 90% Reduce  0%
Map 100% Reduce  0%
Map 100% Reduce 33%
Map 100% Reduce 67%
Map 100% Reduce 100%
Elapsed time is 112.558576 seconds.
```

Fig.12 MapReduce Progress for 3 Key-Value pairs

Fig.12 shows Map tasks and 3 reduce tasks compared to fig. 4 which shows only one reduce tasks. The fig shows little performance improvement of MapReduce . The number of map tasks is always greater than the no of reduce tasks

```
Trial>> tbl=readall(MaxBlue);
Trial>> disp(tbl);
        Key                    Value
    'Max Red Color Image'      'C:\Users\user\Dropbox (KEC)\red.jpg'
    'Max Green Color Image'    'C:\Users\user\Dropbox (KEC)\Color-Green.jpg'
    'Max Blue Color Image'     'C:\Users\user\Dropbox (KEC)\2015-04-03 14.14.22.png'
fx Trial>>
```

Fig.13 Display of Key-values for Red, Green and Blue

Fig.13 shows three keys and its associated values. The location of maximum Red, Green and Blue color is displayed under value field.

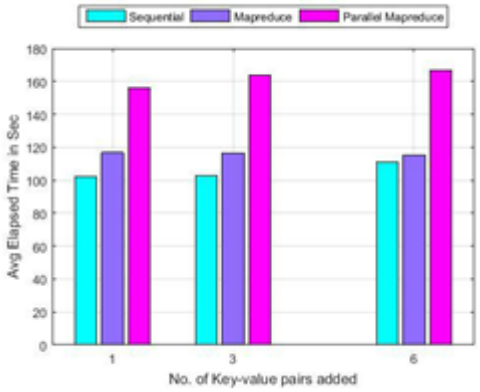

Fig.14 Comparing Performance related to different Key-Values

Fig.14 displays the elapsed time for one, three and six keys. There is considerable performance improvement of MapReduce when no. of keys are increased. The same is not true with Parallel MapReduce.

5.3 Exp #3: Six Key Values and finding maximum of Red,Green, Blue, Yellow, Magenta and Cyan – with dependencies between iterations.

```
Trial>> disp(tbl);
        Key                              Value
    'Max Blue Color Image'               'C:\Users\user\Dropbox (KEC)\2015-04-03 14.14.22.png'
    'Max Green Color Image'              'C:\Users\user\Dropbox (KEC)\Color-Green.jpg'
    'Max Red Blue(Magenta)Color Image'   'C:\Users\user\Dropbox (KEC)\pink.png'
    'Max Red Color Image'                'C:\Users\user\Dropbox (KEC)\red.jpg'
    'Max Red Green(yellow)Color Image'   'C:\Users\user\Dropbox (KEC)\yellow.png'
    'Max Blue Green(cyan)Color Image'    'C:\Users\user\Dropbox (KEC)\cyan.png'
fx Trial>>
```

Fig.15 Displaying Key-Values for Red, Green, Blue, Magenta,Cyan and Yellow

Fig. 15 Result of 6 color key-values.. The location of images with maximum Red, Blue, Green, Yellow, Magenta and Cyan are displayed. The results will be displayed in any order i.e results are not sorted.

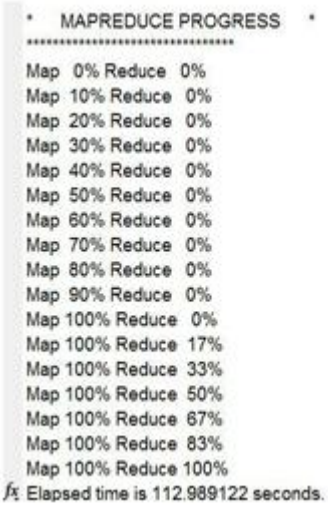

Fig. 16 MapReduce Progress for 6 Key-Value Pairs

The Fig.16 Shows 10 map tasks and 6 reduce tasks. This brings a very little improvement in performance of MapReduce. Still there is no considerable improvement in Parallel MapReduce due to the dependency of code between iterations.

5.4 Exp #4: Six Key Values and finding maximum of Red,Green, Blue, Yellow, Magenta and Cyan – No Dependency between iterations.

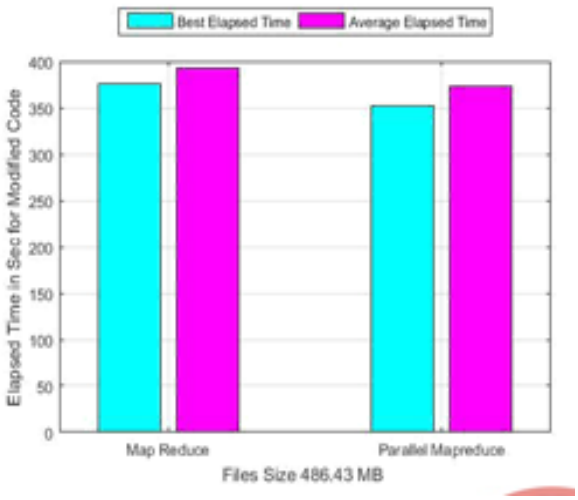

Fig.17 Comparing Performance improvement related to Coding Style with File Size170.34MB

Fig. 17 shows that there is improvement in MapReduce when there is no dependencies of values betweeen iterations. Sequential and MapReduce shows equal performance. But there is no expected improvement in Parallel MapReduce. There is performance improvement when the for loop in sequential code is executed parallely.

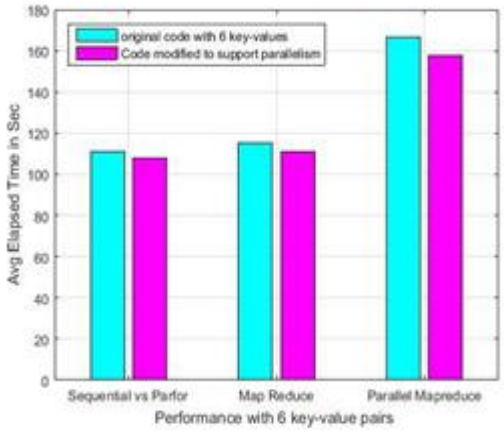

Fig.18 Performance of MapReduce and Parallel MapReduce with improvized code and File Size

Fig.18 displays the expected improvement of Parallel MapReduce. With more Reduce functions and more FileSize there is improved performance. High performance could not be achieved due to start up latency [20] for worker processes. Unexpected high elapsed time happens at times but the program need not be started from scratch [20]

5.5 Exp #5: Six Key Values and finding maximum of Red,Green, Blue, Yellow, Magenta and Cyan – No Dependency between iterations involving multiple workers

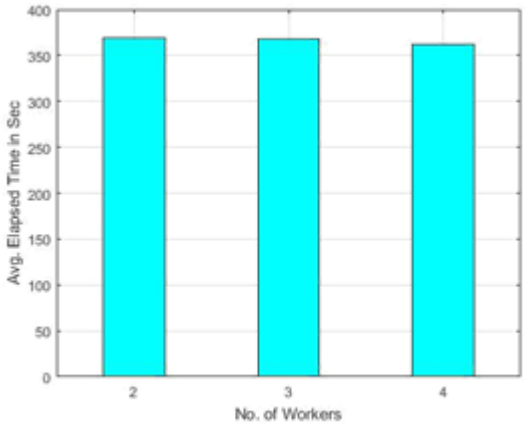

Fig.19 Involving More Number of Workers

Fig.19 Experiments with more than two workers. If the system uses dual core by default it is possible to use two workers only. The experiment can be extended with 5 workers also with no appreciable result. By using more no of workers there is less improvement in average elapsed time because of more start up latency involved. Matlab supports upto 512 workers.

VI. CONCLUSION

The efficiency of MapReduce and Parallel MapReduce lies mainly with the efficiency of the code. Large volume of data constitutes for speedup in parallel MapReduce. Adding more

workers helps in little speed up. DropBox retains the Quality of the image. Even if multiple cores are available it will not be of much use if sufficient RAM capacity is unavailable [12]. Sufficient memory , ample number of cores that support mpre workers and perfect coding style with large file size will help in speedup. Communication cost is minimized by use of cores.

VII. FUTURE WORK

Graphics processing Units (GPU) can be used instead of Central processing Units (CPU) to improve the speed of parallel processing of images. GPU has lot of cores and processing is speedy [11]. Recent laptops come with Nvidia graphic card and it can be used for parallel image processing. But the Architecture of GPU is different from CPU and codlings have to be made accordingly. Moreover the same problem can be executed using Quad core processors supporting hyper threading. MapReduce can be implemented using Amazon EC2 cloud services. Support for rich applications should be extended by key-value stores [13]**.**The research can be extended to video data also which is used more in recent days

REFERENCES

1. https://www.dropbox.com/(accessed on 9.5.2016)
2. https://www.flickr.com/(accessed on 9.5.2016)
3. (accessed on 20.4.2016)
4. https://get.hike.in/(accessed on 9.5.2016)
5. https://photos.google.com(accessed on 9.5.2016)
6. http://in.mathworks.com/help/matlab/mapreduce.html(accessed on 19.4.2016)
7. http://in.mathworks.com/help/matlab/large-files-and-big-data.html(accessed on 19.4.2016)
8. http://in.mathworks.com/help/distcomp/program-independent-jobs-on-a-local-cluster.html(accessed on20.4.16)
9. https://www.whatsapp.com/(accessed on 9.5.2016)
10. https://picasaweb.google.com/home(accessed on 9.5.2016)
11. Bingsheng He, Wenbin Fang, Naga K. Govindaraju, Qiong Luo and Tuyong Wang, "Mars: A MapReduce framework on graphics processors, Proceedings of the 17th international conference on Parallel architectures and compilation techniques in ACM, (2008) 260-269.
12. Diana Moise, Denis Shestakov, Gylfi Gudmundsson, Laurent Amsaleg. Indexing and Searching 100M Images with Map-Reduce. Proceedings of the ACM International Conference on Multimedia Retrieval, (2013) 17-24.
13. Divyakant Agrawal, Sudipto Das, Amr El Abbadi, Big Data and cloud computing : Current state and future opportunities, Proceedings of the 14th International Conference on Extending Database Technology in ACM , (2011) 530-533
14. Domenico Talia,"Clouds for Scalable Big Data Analytics", IEEE Computer Society, (2013) 98-101 [15]Fabrizio Marozzo, Domenico Talia, Paolo Trunfio, P2P – MapReduce : Parallel data processing in dynamic Cloud environments, Journal of Computer and System Sciences, Volume 78, Issue 5, (2012)1382-1402
15. Haluk Demirkan, Dursun Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud", Decision support systems in Elsevier, volume 55, issue 1 (2013) 412-421.
16. Hyeokju Lee, Myoungjin Kim, Joon He`r and Hanku Lee, Implementation of MapReduce-based Image Conversion Module in Cloud Computing Environment, IEEE international Conference on Information Network, (2012) 234-238
17. Jaliya Ekanayake, Hui Li, Bingjing Zhang, Thilina Gunarathne, Seung-Hee Bae, Judy Qiu, Geoffrey Fox, Twister: a runtime for iterative MapReduce, Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, (2010) 810-81
18. Jeffrey Dean, Sanjay Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, Communications of the ACM, Vol. 51, No.1 (2008) 107-113
19. Jeffrey Dean, Sanjay Ghemawat, MapReduce: A Flexible Data Processing Tool, Vol. 53, No. 1, and Communications of the ACM, (2010) 72-77
20. C. Ji , Y. Li , W. Qiu , U. Awada , K. Li," Big Data Processing in Cloud Computing Environments", IEEE 12th International Symposium on Pervasive Systems, Algorithms and Networks, (2012) 17-23
21. D. Jiang, B. Ooi, L. Shi, and S. Wu, The performance of mapreduce: An in-depth study, Proceedings of the VLDB Endowment, Vol. 3, No. 1-2 (2010) 472-483.
22. Junbo Zhang, Dong Xiang, Tianrui Li, Yi Pan, M2M: A Simple Matlab-to-MapReduce Translator for Cloud Computing, Tsinghua Science and Technology, Vol 18, No.1 (2013) 1-9
23. D. Kossmann, T. Kraska, and S. Loesing, An evaluation of alternative architectures for transaction processing in the cloud, Proceedings of the ACM international conference on Management of data., 2010, pp. 579–590
24. Michael Stonebraker, Daniel Abadi, David J. DeWitt, Sam Madden, Erik Paulson, Andrew Pavlo , Alexander Rasin, MapReduce and parallel DBMSs: friends or foes? , Communications of the ACM, Volume 53, Issue 1 (2010) 64-71
25. Raja Appuswamy ,Christos Gkantsidis, Dushyanth Narayanan, Orion Hodson, Antony Rowstron , Nobody ever got fired for buying a cluster, Technical Report of Microsoft Corporation, (2013). http://www.research.microsoft.com
26. C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, C. Kozyrakis, Evaluating mapreduce for multi-core and multiprocessor systems. Proceedings of 13th International Symposium on High-Performance Computer Architecture(HPCA), (2007) 13-24
27. Richard McCreadie, Craig Macdonald, Ladh Ounis, MapReduce indexing strategies: Studying Ounis, MapReduce indexing strategies: Studying Ounis, MapReduce indexing strategies: Studying Management, Volume 48, Issue 5, (2012), 873-888.
28. Steven J. Plimpton, Karen D. Devine, MapReduce in MPI for Large-scale graph algorithms, Special issue of Parallel computing, Volume 37, Issue 9, (2011), pp.610-632
29. K. Wiley, A. Connolly, J. Gardner, S. Krughoff, M. BalaZinska, B. Howe, Y.Kwon, Y. Bu, Astronomy in the Cloud: Using MapReduce for Image Co-Addition, The Astronomical Society of the Pacific, Volume 123, Number 901, (2011)
30. Yuzhong Yan, Lei Huang, Large-Scale Image Processing Research Cloud, The fifth International Conference on Cloud Computing GRIDs and Virtualization, 2014, 88-93
31. C. Zhang, F. Li, J. Jestes, Efficient parallel kNN joins for large data in mapreduce, Proceedings of the 15th International Conference on Extending Database Technology in ACM, (2012) 38-49.

AUTHORS PROFILE

Buvaneswari V.B presently working as Assistant Professor in Government Arts College, Coimbatore, TamilNadu. She has more than 14 years of teaching experience in teaching in India. Received M.Sc in Computer Science in 2000 at Periyar University. Received M.E in Computer Science in 2007 at Anna University Chennai and currently pursuing Ph.D at Anna University Chennai. Her primary research interest includes Service Oriented Architecture, Network Security, Cloud Computing, and Big Data Analytics. And Image processing

Dr. S. Shanthi received her PhD degree in Computer Science and Engineering at Anna University, Chennai, India in 2015. She is presently working as an Assistant Professor (SLG) in the Department of Computer Applications, Kongu Engineering College, Tamil Nadu, and India. Her area of interest includes, Data Mining, Image Processing, Pattern Recognition, Big data analytics, Health care Informatics and Soft Computing.

M.Pyingkodi is an Assistant Professor in the Department of Computer Applications, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India. She received her Bachelor's degree in Computer Science at 2003 and a Master's degree in Computer Applications from Bharathiar University at 2006. Her areas of specializations are Data Mining in Bioinformatics, having teaching experience around eleven years. She is pursuing her Ph.D in Computer Science at Anna University, Chennai.