RESEARCH ARTICLE OPEN ACCESS

Image Processing and Handover Techniques for Map Reduce Based on Big Data

Revathi Durgam
Dept. of computer science & Engineering
AVN institute of technology,Hyderabad

vijendar amgothu
Dept. of computer science & Engineering
University College of Engineering
Osmania University,Hyderabad

Abstract:

Cloud computing is the one of the emerging techniques to process the big data. Large collection of set or large volume of data is known as big data. Processing of big data (MRI images and DICOM images) normally takes more time compare with other data. The main tasks such as handling big data can be solved by using the concepts of hadoop. Enhancing the hadoop concept it will help the user to process the large set of images or data. The Advanced Hadoop Distributed File System (AHDF) and MapReduce are the two default main functions which are used to enhance hadoop. HDF method is a hadoop file storing system, which is used for storing and retrieving the data. MapReduce is the combinations of two functions namely maps and reduce. Map is the process of splitting the inputs and reduce is the process of integrating the output of map's input. Recently, in medical fields the experienced problems like machine failure and fault tolerance while processing the result for the scanned data. A unique optimized time scheduling algorithm, called Advanced Dynamic Handover Reduce Function (ADHRF) algorithm is introduced in the reduce function. Enhancement of hadoop and cloud introduction of ADHRF helps to overcome the processing risks, to get optimized result with less waiting time and reduction in error percentage of the output image.

Keywords: Cloud computing, big data, AHDFS, map reduce, ADHRF algorithm.

I. Introduction

Cloud computing is the required after field nowadays in information technology. Cloud computing is a package comprising of server and client machines. Cloud computing processes the data in the distributed and parallel modes. Cloud computing is also, known as service on demand. The services of the cloud computing enables end users to pay and obtain required data from the service providers like IBM, AMAZON and INTEL among others. In this proposed work, an enhanced cloud tool called INTEL (a product of Intel) is utilised. Enhancing the concept of hadoop over the cloud computing gives the better result in the process of computing big data. The hadoop enhances the Advanced Hadoop Distributed File System (AHDFS) and MapReduce functions in it. The MapReduce concept will execute the complicated tasks very easily, with simple requirements of machines. Google first introduced the concept of MapReduce programming model [2, 6]. MapReduce concept has few basic functions like master, slave, job manager, job node, etc. The master function supervises the execution of map and reduces operations.

The image processing techniques like grayscale, sobel edge detection, gaussian blur and fast corner detection are also, enhanced in the proposed work. Presently, this regular set of work is made with the other corner detection method and scheduling algorithm for 2D to 3D data processing [2, 11]. In the proposed work, it has been proved that, there is another better corner method, improved Sum of

Absolute Differences (SAD) matching and an optimized scheduling algorithm, which could benefit the client in the useful manner. Advanced Dynamic Handover Reduce Function (ADHRF) algorithm has proved that, it works better than the existing algorithm in the reduce function. JPEG files can be viewed and opened mostly by many image viewers. Some image formats may get deleted while compressing. Even some reduction in the quality of image may occur while compressing. In the proposed model, a template has been made in such a way that, it accepts the input data in any format. An attempt has been made to show that all the accepted input is to be compressed to that of the fixed frame size. The raw data formats are converted to the fixed frame size and then the data compression is done to a fixed scale. The output will be a better one with high flexibility, less waiting time and less error percentage. Mostly the medical data will be in the DICOM format and rarely in the .JPEG format. As an outcome of this work the output templates received will be in the .JPEG format. Implementation of fast corner method has proved That it can give the users a better result than harris corner method. Also, the implementation of improved SAD reduces the error percentage when compared to the existing method. The computation of small files is proved to be better in the existing system. The concept of big-data is being used in the proposed work, so that it could manage the input into fixed frame size. Patel *et al.* [8] reports the experimental work on big data. The recent problem and its optimal solution using hadoop cluster.

AHDFS for storage and using parallel processing to process large data sets using map reduce programming template. In the reported work, the word junks meaning intermediate data will be sent to the reduce function. In the reduce function, the proposed ADHRF algorithm is added to give the result. Big data chunks with different size and sequence is 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 [3, 4, 5]. These junks are computed in the reduce functions. The data transfer delay can be comparable or even higher than the time required computing the data [1, 9, and 10]. To overcome the problem of transfer delay from the existing system, a novel optimization scheduling algorithm has been implemented in our study. In the existing module, processing the large data by a large number of small files, which exhibits better performance [7]. In this Work, handling of large set of data has been implemented with the enhanced tool.

The Figure 1 shows format by which the inputs in the various formats are stored in the data container. The job manager function is use to assign work for the servers. These inputs of the data will be stored in AHDFS to start the map reduce function. Each job will be assigned as a task.

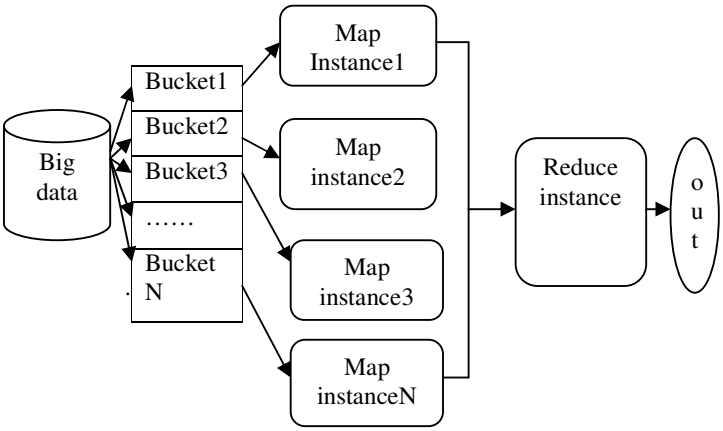

Fig.1.Architecture of hadoop map reduce method

The hadoop's method is to split input data and distribute to the hosts to compute. This work will be done simultaneously in the parallel mode. This is known as distributed cum parallel computing. Figure 2 shows the modules in the map reduce function. The task depends upon the strength and the storage of the computing system.

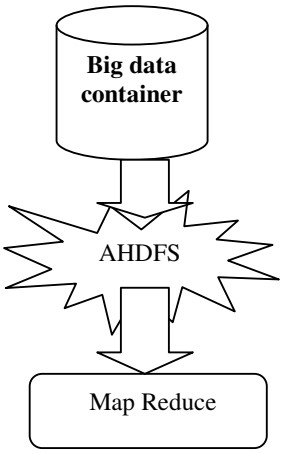

Fig.2. Map reduce modules

II. Proposed Method

The proposed method contains some of the functions. That is the big data processing ,mapreduce and AHDF.

II.1. Big Data Processing method

In this work, the process of comprises of few data Processing techniques as shown in Figure 3. Such as grayscale, sobel, gaussian blur, fast corner and SAD matching to find the difference between two data. The grayscale method is used because, when an image is converted to grayscale, the image's quality will be improved. Sobel method is used to find the edges of the images. The gaussian blur is used to blur the image, so that it will be useful for fast corner_9 method to detect the corners of the image. The enhancement of these methods results in better quality of the image. Hence, these are surveyed as the best. The given equation finds the solution for the harris corner method to find the image patch area with the argument (u, v). Where (u, v) denotes the image patch point and while processing, (x, y) get shifted to (u, v), where w is the center point on the (x, y).

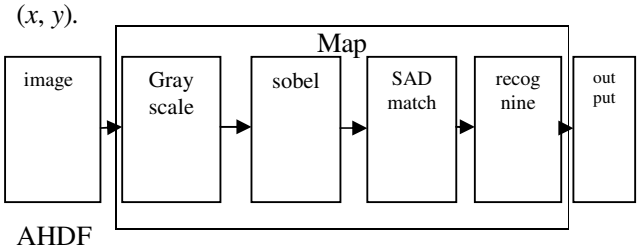

Fig.3.Data processing method

Hadoop is an Apache open source framework which is written in java. High volumes of data, in any structure, are processed by Hadoop. Hadoop allows distributed storage and distributed processing for very large data sets. The main components of Hadoop are:

1. Hadoop distributed file system (HDFS)
2. MapReduce

Hadoop has three layers. The two major layers are MapReduce and AHDFS.

AHDFS (Storage layer):- In the proposed method the Hadoop techniques has a Advanced distributed File System called AHDFS, which stands for Advanced Hadoop Distributed File System. It is a File System used for storing very large files with streaming data access patterns, running on clusters on commodity hardware [8].There are two types of nodes in AHDFS cluster, namely name node and data nodes. The name node manages the file system name space, maintains the file system tree and the metadata for all the files and directories in the tree. The data node stores and retrieve blocks as per the instructions of clients or the name node. The data retrieved is reported back to the name node with lists of blocks that they are storing data. Without the name node it is not possible to access the file. So it becomes very important to make name node resilient to failure 11].

MapReduce (Processing/Computation layer):- It is a function which is meant for managing applications on multiple distributed servers. In MapReduce function is dividing and conquers method is used to break the large complex data into small units and process them. It reads the data from AHDFS in an optimal way. However, it can read the input data from other places too; including mounted local file systems, the web, and databases. It divides the computations between different computers (servers, or nodes), is also called fault-tolerant. If some of nodes fail, Hadoop knows how to continue with the computation, by re-assigning the incomplete work to another node and cleaning up after the node that could not complete its task. It also knows how to combine the results of the computation in one place[9].The other core components in Hadoop architecture includes Hadoop YARN, it is a framework for job scheduling and cluster resource management. The other component is the cluster which is the set of host machines (nodes).

II.2. Methodology

The proposed method is the ADHRF algorithm. After the installation of INTEL Manager is over, next the hadoop set up has to be done in the system. Now the system can work on the MapReduce functions and use the facility of the HDFS. After the successful installation of the INTEL Manager and hadoop and its content, the proposed DHRF algorithm has to be inserted in the reduce function. Since the MapReduce is an open source, it can be edited and modified according to the user's need. When a task is applied to the nodes on the cluster, the map function starts its job of splitting the data. The task node assigns the job for the each node, and also it supervises the job node and its functions. When the job assigned by the task node gets over, the

output of the map function is ready with the intermediate data. In the proposed work to obtain an optimized result from the existing image processing techniques, the proposed work implements the hadoop and cloud computing using INTEL Manager. A set up of ten machines configured with Intel (R) Core 2 duo, 4 GB RAM and 2.93 GHz processor is used in this study. In these machines, INTEL Manager a Cloud tool is installed. INTEL Manager has basically Master and Slave in it. Since hadoop is an open source, it can be modified according to our need. So, the editing is done in the reduce function, by adding the proposed DHRF algorithm. In this work, an enhanced cloud tool called INTEL Manager is utilized. The coding or the application of image processing techniques is installed on the INTEL.Manager, to run the proposed experiment. The main advantage of this enhanced INTEL Manager is it works on the Normal Window XP (64 bit) infrastructure. In this work INTEL Manager which is open source software available as private cloud cum hybrid cloud is used.

III. Conclusions

This work proves the images of various formats can be taken as input. The quality of the image is fine tuned with the proposed algorithm which has produced better result in the .jpeg format. This result shows that, whatever may be the format of the input, the result can be obtained in .jpeg format to give better improved quality of output with less waiting time and error percentage. In this work, the present algorithm has been implemented for the optimizing the result in the reduce function. In future, it is planned to incorporate some more modification on the map function, so that the results can be more accurate. This work implements four image processing techniques, where as in the future work, the comparison testing can be done by using less number of image processing techniques.

IV. References

[1] Chen C., Wu T., Kuo L., Lee W., and Chao H., "Cloud-Based Image Processing System with Priority-based Data Distribution Mechanism," *Computer Communications*, vol. 35, no. 15, pp. 1809-1818, 2012.

[2] Gao K., Wang Q., and Xi L., "Reduct Algorithm based Execution Times Prediction in Knowledge Discovery Cloud Computing Environment," *the International Arab Journal of Information Technology*, 2013.

[3] HaoY., "Fast Corner Detection-machine Learning for High Speed Corner Detection," 2010.

[4] Jung G., Gnanasambandam N., Mukherjee T., "Synchronous Parallel Processing of Big-data Analytics Services to Optimize Performance in Federated Clouds, Cloud Computing (CLOUD)," *in Proceedings of the 5th International Conference on Digital Object Identifier*, pp. 811- 818, 2012.

[5] Kalagiakos P. and Karampelas P., "Cloud Computing learning, Application of Information and Communication Technologies (AICT)," *in Proceedings of the 5th International Conference on Digital Object Identifier*, pp. 1-4, 2011.

[6] Lee H., Kim M., Her J., and Lee H., "Implementation of Mapreduce-based Image Conversion Module in Cloud Computing Environment," *in Proceedings of the International Conference on Information Network*, pp. 234-238, 2012.

[7] Malik S., Huet F., and Caromel D., "Cooperative Cloud Computing in Research and Academic Environment using Virtual Cloud, Emerging Technologies (ICET)," *in Proceedings of the International Conference on Digital Object Identifier*, 2012.

[8] J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S.-H. Bae, J. Qiu, and G. Fox, "Twister: A runtime for iterative mapreduce," in Proc. 19th ACM Symp. High Performance Distributed Comput., 2010, pp. 810–818.

[9] D. Peng and F. Dabek, "Large-scale incremental processing using distributed transactions and notifications," in Proc. 9th USENIX Conf. Oper. Syst. Des. Implementation, 2010, pp. 1–15. [10] D.

[10] Logothetis, C. Olston, B. Reed, K. C. Webb, and K. Yocum, "Stateful bulk processing for incremental analytics," in Proc. 1st ACM Symp. Cloud Comput., 2010, pp. 51–62.

[11] J. Cho and H. Garcia-Molina, "The evolution of the web and implications for an incremental crawler," in Proc. 26th Int. Conf. Very Large Data Bases, 2000, pp. 200–209.