IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 65-68
www.iosrjournals.org

A Review on Hadoop MapReduce using image processing and cloud computing

Minakshi M. sonawane[1], Santosh D. Pandure[2], Seema S. Kawthekar[3]
[1](Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada university, Aurangabad, India)
[2](Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada university, Aurangabad, India)
[3](Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada university, Aurangabad, India)

Abstract: *Hadoop is an open source framework that allows distributed processing of large data set across clusters of computers. Big data describes technology to capture, store, distribute and manage the large size data set. Data is generated by different sources such as from a web site or click streams (e.g. net ix, Face book, Google), Sensors (energy monitoring, application monitoring, telescope) and biomedical diagnosis. Image processing is perform important function in various research areas such as biomedical imaging, remote sensing, astronomy, internet etc. Hadoop image processing library is used with the apache Hadoop map reduce programming framework. Hadoop image processing provide facility to high- throughput image processing. Splitting digital image in to multiple segments using image segmentation technique which is used to identify the object and boundary of object in image processing. Images are stored in Hadoop distributed file system after that apply map reduce algorithm to extract the features from images. Reducer collect all result and combines from all map function and stores in the result of HDFS. Hadoop image processing support to cloud computing which is type of Internet based computing that provides shared computer processing resources and data to computers and other devices on demand.*

Keywords: *Hadoop, Big data, Image processing, Map reduce, HDFS, Cloud computing.*

I. Introduction

Hadoop is an open source framework which is used to processing on large data sets. Many times hardware failures are occur and it will be handled by Hadoop framework. Hadoop is consist a storage part known as HDFS and processing part known as Map reduce. The HDFS (Hadoop distributed File System) is a scalable, distributed and portable file system which is written in java. HDFS contain large files across multiple machines. Map Reduce framework consist of Map and Reduce, map method performed filtering and sorting of data and Reduce method performs summary operations. Hadoop framework is composed of the different modules such as Hadoop common contain all libraries which is required. HDFS library hold HDFS store data on machine. Hadoop YARN is resource management platform which manage computing resources. Hadoop Map Reduce is implement map reduce programming model for large data processing. Hadoop image processing interface is the library which is used to perform the image processing using map reduce framework. Hadoop image processing library provide solution to store large amount of images on Hadoop distributed file system. This library provide the implementation with opencv (open source computer vision library).Hadoop working with different demeans such as name-node is run on master node. Data-node is runs on Slave-node. The Name-node instructs data files to be split into blocks, each of which are replicated three times and stored on machines across the cluster. Client machine is responsible for loading data into the cluster. It will submit Map Reduce jobs and viewing the results of the jobs. Job-tracker tracks jobs which split into cluster. The Task-tracker accepts tasks from the Job-Tracker. The reduce phase extract image stored in HDFS the specific processing is show in figure 1.

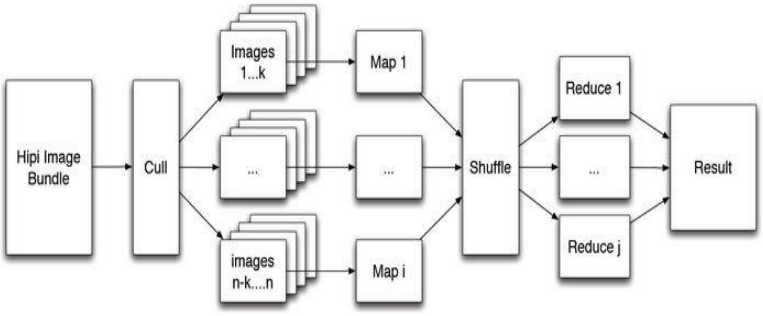

Fig 1. Hadoop image processing interface

Hadoop image processing interface library designed is used with parallel programming framework.it provide how to storage a large collection of images on Hadoop distributed file system and make available or efficient distributed processing and it is integrated with opencv a popular open source library. The HIP Image is base class provide to the underlying grid of pixel image value as array , bytes and floats, respectively.it provide a number of useful function like crop, color space and conversion and scaling.

1. **Hipi Image Bundle:** It is open source framework. It maintained by a group of dedicated researchers and developers.
2. **Cull:** The initial stage of a Hadoop image processing interface program is a culling step that allows filtering the images in a Hadoop based on a user denied condition are used in spatial resolution and criteria related to the big data.
3. **Images:** The primary presentation for a collection of images on the Hadoop distributed file system. Map reduce is optimized to support efficient processing of large file system. HIB is actually compared two file stored on the system.
4. **Shuffle:** shuffle can start before and after map phase has finished to save same time. Reduce status is greater than 0 percent but less than 3% when map status is not yet 100%.
5. **Mapper:** Take the data and convert into another set of data where Individual elements are broken down into tuples.
6. **Reducer:** Take the data from mapper and combine those data tuples into smaller set of tuples.

II. Related Work

With the vast growth of images produce by 2D/3D graphical hardware technology .the task of image retrieval analysis framework identify and it provide data as well as map reduce phase to big no of medical data. Hadoop is an open source framework which is handle big data and it provide scalability fault, tolerance, high availability and parallelism. It can handle vast amount of data it manage such as master node it can handle map reduce task, salve node and prototype task it find query image, map reduce task given tired retrieval system. It is stored in Hadoop distributed file system. Job tracker are capable to handle smallest task which is come first [1]. The evolution of digital technology and stored device it is fast development of digital image library, and all kinds digital tools produce number of image every day. It becomes a warm research in recent years. The content-based medical image retrieval (CBMIR) speed, and high precision. It has been extensively applied in the fields such as medical, aided medical diagnosing, and medical information management [2]. Cloud computing can distributed task to every node in parallel processing it provide research idea for medical image retrieval [3]. There are two main aspects of development of parallel image processing system; one is algorithms. It is searching the efficient parallel algorithm and development of high-performance parallel computer to achieve specific purposes, but such system is limited to the scope of application. The other is developed for general-purpose parallel image processing system, which is the main stream of the parallel image processing system [4]. It is cannot be described the word, it only understanding different image from person to person it applied in such as medical technique medical diagnosing, medical information management with the support next generation like big data analytics, one can find and improve the medical image processing. Hadoop framework is one of the finding based on map reduce distributed computing model. It most widely used in parallel computing is designed to scale from one to multiple machines, each are stored in computing that is divided into based method. We used to map reduce computing model to extract feature files into Hadoop based open source distributed, and column oriented store model in big data. Split the digital image using image segmentation for detection of object and its boundaries. Every pixel value can be extract with the label and it will give the visual characteristics [5]. The Name Node is responsible for managing the name space of the file system and the access of the clients to the files, while Data Node manages the storage of the data of its node, handles the client's reading and writing requests of the file system, as well as carries on the creation, deletion and copy of the data block under the unified scheduling Name Node [6].

III. Different Techniques

Following table 1 describe the various techniques and result of different research papers.

Sr.No	Title	Author and Publication	Year	Methodology	Results
1	Large-Scale Image Processing Research Cloud	Yuzhong yan, Lei Huang	2015	DFT Fourier Transform, Spital Domain, Frequency Domain, Digital many image file, openCv, HPC Custer.	Big size of images in small HPC cluster with big size of 1m to mb group of image [7].
2	Comparing apache spark and map reduce with performance analysis	Satish gopalni,rohan arora	March 2015	Big data, machine learning, k means.	The data set decrease processing time of compared mapreduce

	using k means				[8].
3	Massive Medical Image Retrieval System Based On Hadoop	YAO Qing-An,ZHENG Hong, XUZhong-yu,WUQiong ,LI Zi-Wei, YunLifen .	February2014	Brushlet Transform; Local Binary Patterns; Distributed System ,SIFT Scale Invariant Feature	hadoop is based on retrieval system can reduce time of storage and improve image retrieval speed[9].
4	HIPI: A Hadoop Image Processing Interface for Image-based MapReduce Tasks	Sweeney, C. Liu	2011	Hadoop image processing interface	Creating large scale vision applications [10].
5	Comparative Study of Different Methods for Brain Tumour Extraction from MRI Images using Image Processing.	Neha Baraiya , Hardik Modi	January 2016	MRI, Morphological Operations, Thresholding, Tumour Extraction ,Image Registration, Image Pre-processing, Segmentation Techniques	Accuracy of the thresholding and watershed are 87.48, 9134, 92.76 high contrast on tumor and normal part [11].
6	MULTIDIMENSIONAL TEXTURE CHARACTERIZATION: ON ANALYSIS FOR BRAIN TUMOUR TISSUES USING MRS AND MRI	Deepasubramaniam Nachimuthu, runadevi Baladhandapani	August-2014	Magnetic resonance spectroscopy, Magnetic resonance imaging, Multidimensional co-occurrence matrices, Feature extraction, Extreme learning machine, Particle swarm optimization.	Overall distributed accuracy of 86.5% both of volumetric and spectroscopic feature provide the highest discrimination accuracy between low to high gliomas in 99.15% [12].
7	Enhancing The Efficiency of Parallel Genetic Algorithms for Medical Image Processing with Hadoop	D. Peter Augustine	December 2014	Parallel genetic algorithms, Health care, HealthCare Applications, Hadoop, and Cloud Computing.	The study of medical image processing that to narrowed to the brain image with the parallel genetic algorithms in the cloud environment [13].
8	Map Reduce Framework Implementation on the Prescriptive Analytics of Health Industry	Lalit Malik, Sunita Sangwan	June 2015	HDFS, DFS, Hive, Pig	The move data out of storage system while parallelizing the computation, the due to increasing number of sensor and resulting data [14].

Table No. 1: different techniques in Hadoop image processing

IV. Conclusion

Hadoop image processing library support to store and retrieve the large volume of image bundle with Hadoop distributed file system. Large volume of visual data is acquired by biomedical imaging, remote sensing, astronomy, Internet and their need for efficient and effective processing simulate the use of Hadoop image processing. Hadoop image processing is use to medical diagnosis such as brain tumor detection, iris detection, face recognition etc. Hadoop image processing frame work facilitates efficient and high-throughput image processing with MapReduce frame work. Hadoop base, cloud computing, image segmentation, thresholding and image enhancement technique are used in Hadoop image processing.

Acknowledgements

I take this opportunity to sincerely acknowledge the University of Grant Commission (UGC), Government of India, New Delhi, for providing financial assistance in the form of Rajiv Gandhi National Fellowship which supported me to perform my work comfortably. Also Authors gratefully acknowledge support by the Department of computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India.

References

[1] Jyoti S.Patil,G.Pradeepani,"Two Dimensional Medical Images Diagnosis using MapReduce",*pp 1-5, Vol 9 (17), May 2016*.

[2] Wang Zhongye,Yang Xiaohui,Niu Hongjuan",Brushlet domainretrival algorithum based on complex computer simulation of image texture characteristics, *pp 263-266 ,2011*.

[3] ZHANGJ,LIUXL,LUOWJW,BOLTN,"Distributed image retrival system based on mapreduce", *pp 93-98, 2010*.

[4] Nilesh lohar, dipankar chavan, sanjay arade, amol jadhav,deepti chikmurge,"Content Based Image Retrieval System over Hadoop Using MapReduce ",*Vol 2, Issue 1, pp. 123-125, January-February 2016.*

[5] Fay C,JEFFEREY D,SANYJAY G,"A distributed storage system or structured data proceedings The symposium on operating system design and implement", *pp 276-290, 2006.*

[6] R.Saraswathy,P. Priyadharshini,P. Sandeepa," Hbase Cloud Research Architecture for Large Scale Image Processing",*vol 4, Issue 12, December 2014.*

[7] Yuzhong yan,Lei Huang,"large-scale image processing research cloud",*pp 88-93, ISBN: 978-1-61208-338-4,2014.*

[8] Gopalani, Satish; Arora, Rohan,"Comparing apache spark and map reduce with performance analysis *using K-Means" International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 1, March 2015*.

[9] Yao Qing-An,ZHENG Hong,XUZhong-Yu, WU Qiong, LIZI-wei,Yun Lifen",Massive medical Images Retrival System Based on Hadoop," *pp 216-222,Vol 9, Issue No 2,February 2014.*

[10] Sweeney, C. Liu, L., Arietta, S., Lawrence, J. HIPI: A Hadoop Image Processing Interface for Image-based MapReduce Tasks. B.s. thesis, University of Virginia (2011)

[11] Neha Baraiya,Hardik",Comparative Study Of Different Methods for Brain Tumar Extraction From MRI Images using Image Processing", *pp 1-5,Vol 9(4),Issue 0974-6846, january 2016*.

[12] Deepa subramanian Nachimuthu,Arunadevi Baladhandapani,"Multidimensional texture Characterization on Analysis for Brain tumar Tissues Using MRS and MRI*",Vol 27,Issue 4,pp 496-506,August 2014.*

[13] D. Peter Augustine,"Enhancing the Efficiency of Parallel Genetic Algorithm for Medical Image Processing with Hadoop",*Vol 108, Issue 17, December 2014.*

[14] Lalit malik, sunita sagwan",Mapreduce framework implementation on the prescritptive Analysis of healh Industry",*Vol 4,Issue 6,pp 675-688,june 2015.*