INTERNATIONAL RESEARCH JOURNAL ON ADVANCED SCIENCE HUB
e-ISSN : 2582 - 4376
Open Access
RSP SCIENCE HUB
(The Hub of Research Ideas)
Available online at www.rspsciencehub.com

Special Issue of First International Conference on Advancements in Engineering, Management, Science & Technology (ICEMST 2021)

Investigation into the Implementation of Medical Imaging

Bharath Bhaskar[1], Arjun Kundapur[2]
[1,2]Electronics and Communication Engineering, PES University. Bangalore, India.
bharathbr12@gmail.com[1]

Abstract

In this article, we evaluate and discuss the implications of data analytics and Hadoop on clinical image analysis based on predictive algorithms. Each day, the healthcare industry analyses massive amounts of data. A large number of images were produced by various instruments on the patient in various medical situations. Numerous image processing methods and techniques are being developed in an attempt to obtain the most accurate information from images in order to provide an accurate diagnosis. To achieve maximum performance, both current and constantly evolving big data and hadoop ideas may provide more from image processing. The article examines the importance of big data analytics and hadoop in medical image processing utilizing HIPI and Map reduction inside the future implementation of Big data analytics (BDA) and Map reduction.

Keywords: Image processing, Pattern recognition, Predictive analysis, Map reducing, Big Data analytics.

1. Introduction

The increasing volume of clinical visual information produced on a frequent basis in hospitals demands the utilization of classic healthcare image investigation and classification techniques to determine the best approach. The quantity of images and overall dimension have increased significantly over the last 20 years. Recent advancements in image processing enable healthcare professionals to assist in the diagnosis and classification of critical events across massive image series. On the other perspective, acquiring complex characteristics from large datasets of 3D/4D images necessitates incredibly sophisticated software applications, hardware, and cutting-edge technology. Healthcare images have a variety of modes as well as a high resolution. There are several existing imaging modalities, and current concepts, such as spectral CT, being frequently produced. For even generally utilized imaging modalities, pixels or Voxel precision has improved. For example, diagnostic CT and MRI spatial resolution have attained a sub-millimeter level. Visualization tools and techniques provide extremely accurate and high-quality 3D/4D pictures of anatomical and physiological anatomical structures. However, using those images for efficient evaluation is not a priority issue. Because of the complicated structures of clinical imaging from numerous anatomical organs combined simultaneously. Due to the massive size of datasets, complexity, and variance of anatomical organs, image analysis is widely considered as a complicated task. Because of the image distortion and low contrast, the borders of anatomical structures were imprecise and unconnected. As a consequence, effectively segregating the images to extract the regions of requirement from the remaining datasets could prove a significant challenge. In this research, there are numerous image processing algorithms using various approaches. However, both results and analysis will differ tremendously depending on the particular applications, diagnostic modalities (CT, MRI, etc.), and other factors. The algorithm that works perfectly for one purpose may not function at all for all durations. Distortion, turbulence, and partial volume effects are typical imaging artefacts that can have an impact on the efficiency of the algorithms. The segmentation algorithms face a severe problem due to the variety of requirements. There are currently no such 100% accurate algorithms which provide suitable results for any type of medical database. Image processing will need to go through numerous procedures while considering a variety of external elements in order to produce an accurate result. We present research on Big Data Analytics with respect to mounted medical images and the mapping reduction approach for the bundles of pictures in the distributed environment. This approach simplifies the work of image analysis in order to get an accurate outcome for precise assessment. The HIPI as well as the results obtained by researchers from implementing mapping reduction during image processing were used to improve the study.[1-5].

2. Big Data and Medical Image

- Medical care
- The government's services
- The retail industry
- Manufacturers

Big data may also provide benefits in all of the above-mentioned domains. According to data, if India's healthcare utilized big data productively and efficiently, this could produce more than $300 billion in revenue per year. Healthcare expenditures in India may be decreased by up to 66.66%, which is currently around 75%. The identical beneficial circumstance was observed in other domains as well.

2.1 Variety:

The phrase "data variety" was precisely what it indicated. It is generated using imaging techniques such as X-Ray, MRI, CT scans, PET, functional MRI scans, image formats, and more. It also originates from the numerous technologies which produce images and the various circumstances under which they were obtained.

2.2 Velocity:

Big Data is concerned with the rapidity with which information flows in from resources such as image acquisition devices, networking, and human interaction, including such healthcare professional discussions, among many others. The flow of information for storage and processing is enormous and constant. Such authentic data can assist researchers and healthcare professionals in creating better decisions which provide fundamental analytical benefits.

2.3 Volume:

The volume of information is significant, and information has been accumulated from various sources at all moments. The data size ranges from kilobytes to gigabytes. Computers, networking, and human contact with technologies may all generate data.[6-11].

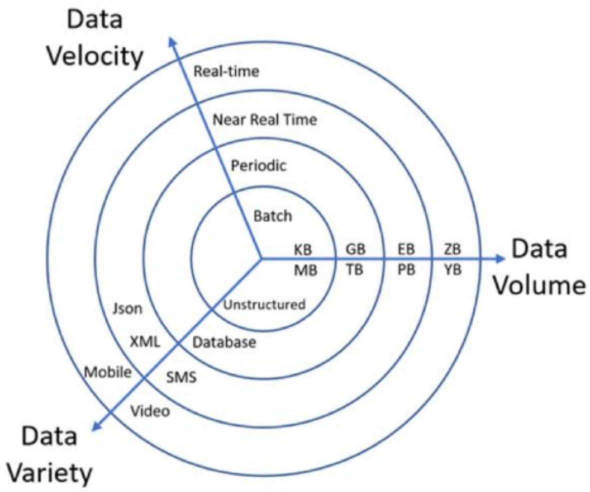


Fig 1: Data 3V

Big Data is a recent major topic in healthcare, both in various fields of study as well as clinical applications. One of the most difficult aspects of medical imaging is finding imaging information in the electronic medical record. Our imaging reports are nearly always unstructured, and our medical pictures are rarely labelled in a way that makes them discoverable or valuable to data mining operations. This must change if medical imaging is to play a significant role in healthcare in this era of big data and customized treatment. Big data plays an important part in diagnostic imaging, and hypotheses aid in the display of pictures and data, as well as diagnosis and therapy. The impact of big data may be considerable inside a variety of domains, and it will have far-reaching consequences for medical imaging as healthcare must monitor, process, optimize, and analyze important patient data. With the above understandings we can affirm that Big Data can be used in several ways, including the following:

- To improve early discovery, analysis, and medical treatment.
- To foresee patient's future health.
- To amplify interoperability and interconnectivity of Healthcare so that the medical professional can gain the needed knowledge from anywhere in the world.
- To enhance patient care by means of remote analysis, remote care and remote medicine by the information gathered from home devices.

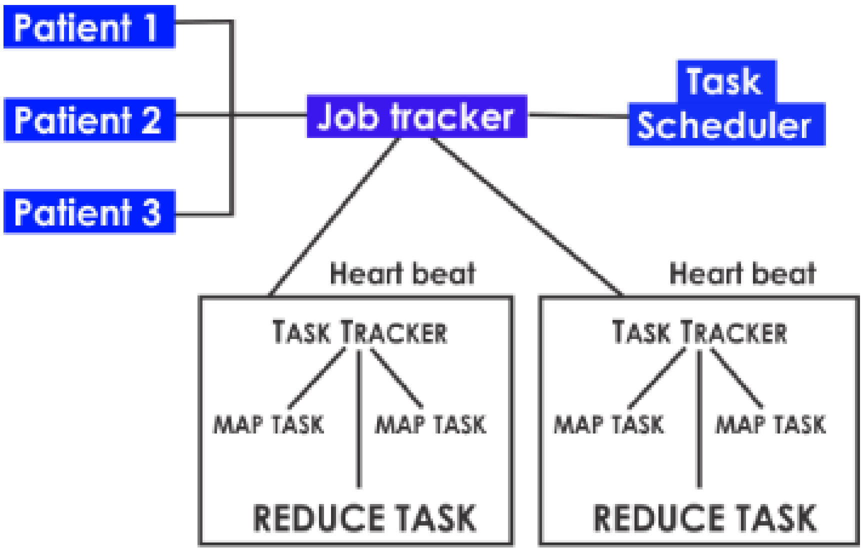


Fig 2 : Map Reduce architecture

3. Software Framework

The Hadoop Cluster computer application would be a framework which incorporates common programming methods that provide to distribute processing of big data collections between clusters of computers. It was designed to extend from the dedicated processor to thousands of computers, within each computation and storage capabilities. Instead of focusing on infrastructure to provide full functionality, the library is designed to determine and handle failures at the application level, enabling a cloud based services to be provided over top of the cluster of servers that may eventually malfunction.[12-18].

Several corporations, such Facebook and Yahoo, are already using architecture of big data processing operations, and can be effectively adapted to operate with either type of hardware, from the single computer to the huge data facility. Hadoop is the best choice for image processing just on MapReduce architecture for its unique features. A Hadoop cluster mainly comprises of a main server one and or even more computational nodes. These computational nodes are already in responsible for storage and computation.

3.1. MAP Reduce

The Clustering algorithm is a cloud based framework which had subsequently been enough to evaluate and explain large-scale images.

Hadoop's MapReduce is a software framework which supports to develop applications to process parallel huge quantity of data like multi-terabyte data sets in a consistent and fault tolerant way on thousands of nodes working together. A Hadoop framework separates an input data set into autonomous large portions, subsequently processed in comparison by predefined mapping processes. The mappings' outputs were categorized by the Hadoop architecture until being providing an input to the reducing tasks. It was all about the operating system. Both the job's input and output must be recorded. The Proposed model automates the process of scheduling tasks, scrutinizing processes, and rerunning tasks while it got failed.

The Hadoop Integrated System Files and the Hadoop architecture can operate on same collection of nodes. The computational and storage nodes are about the same collection. The Hadoop framework effectively distributes odd tasks to nodes where information was previously available as a functional setup. As a result, the cluster's maximum bandwidth is extremely high. Inside the MapReduce system, each cluster node contains a single parent JobTracker and one child Task Tracker. The parent is now in responsible of scheduling tasks, while the child tasks are in responsible of component tasks, inspecting these and re-executing missed work activities. The parent tracker is in responsible of instructing the child.

Because of the efficiency of the Hadoop Framework we can achieve the following benefits for the Image Processing in Healthcare.

Versatile:

To receive the benefits of Hadoop's adaptability, enterprises could easily access multiple data sources and interface to various sets of information (both structured and unstructured).

Performance:

Every backup system of Apache was predicated on such a hadoop cluster which "maps" information to every position on the clusters.

Failure Resistant:

Hadoop is failure tolerant, that's one of its primary strength. Once input could be sent to a particular node, it is also replicated to certain other nodes in the network, providing that in the event of system failure, all backup remains obtainable.

3.2 Medical Image Processing

The field of medicine needs various types of images of a same person in different situations from different devices. The result emerging from the captured images has great impact in the diagnosis. Imaging has become a necessary component in many fields of medical practice to identify, understand and rectify the health problems. X-Ray, MRI, CT scans, PET and functional MRI scans are the instruments for producing various images. Sophisticated computerized quantification and visualization tools are required to analyze of these varied types of images to mine the accurate and investigative result.

The National Electronics Manufacturers Association established the Digital Imaging and Communication in Medicine (DICOM) specification (NEMA). DICOM's challenge is to create clinical images from X-rays, MRIs, CT scans, PET scans, and functional MRI scans allows to share and analyze. The DICOM protocol is indeed a refinement of the NEMA standards. A DICOM file is comprised of two parts: a header and image content. The headers contains data such as the patient's name, the kind of scanning used it to acquire the image, the image's positioning and dimensions, and a host of other characteristics. The image data part comprises any image data.

To reduce disk space the DICOM image data can be compressed either lossless or lossy. Medical Image processing uses real medical images and the supporting environment to demonstrate and explain concepts and to construct perception, imminent and thoughtful.[19-22].

4. Image Based Map Reduce

Conventional Hadoop MapReduce algorithms were capable of effectively managing information input and output. However, they experience issues presenting images inside a way that relates to researchers.

Currently, the methods take extra burden to get the representation of standard float image. For instance, when provide a collection of images to set of Mapping nodes, the user should first provide the

images as a string, then decoding each image for each mapping node before obtaining the image data. It creates extra headache for the users and makes the code messy for understanding and debugging.

HIPI will manage parallelizing procedure and distribute float images to the map function. The images in the HIPI Images Cluster would be spread across all mapping nodes utilizing output specifications with HIPI Image Package. These images were distributed in such a way that the mapping machines as well as the machine where the image lives are as near as possible. For most cases, user had to create Input Format and Record Reader interfaces that determine how the MapReduce task spreads its input and also what information is provided per each server. This is a complex process that causes a lot of discomfort for users. These are handled by user via Input Format and Record Readers. The standard operates using HIPI Image Bundles for a range of image types, sizes, and header information quantities. Each of these various image combinations are handled behind operations in order to achieve floating images to the user. Float images are being sent automatically to the Map operations in an attributed approach. A filtering phase was implemented to the Hadoop workflow during the allocation of inputs and before the map operations begin. Images could be classified based on image properties during the filtering stage. The user defined a sorting group, which Manipulate the images were analyzed. Images that pass the filtering phase would have been transferred to mapping tasks, removing duplication of data. Because the filtering procedure is predicted based on the image header and does not required analyzing the entire image. This method is every efficient one comparing with another techniques.

Images were transmitted as floating images such that users could efficiently obtain pixel values during image processing and vision operations. Regarding storage efficiency, images still are kept in basic image formats (e.g., JPEG, PNG, etc.), while HIPI performs image processing. Inside the Hadoop workflow that provide the user with floating images.

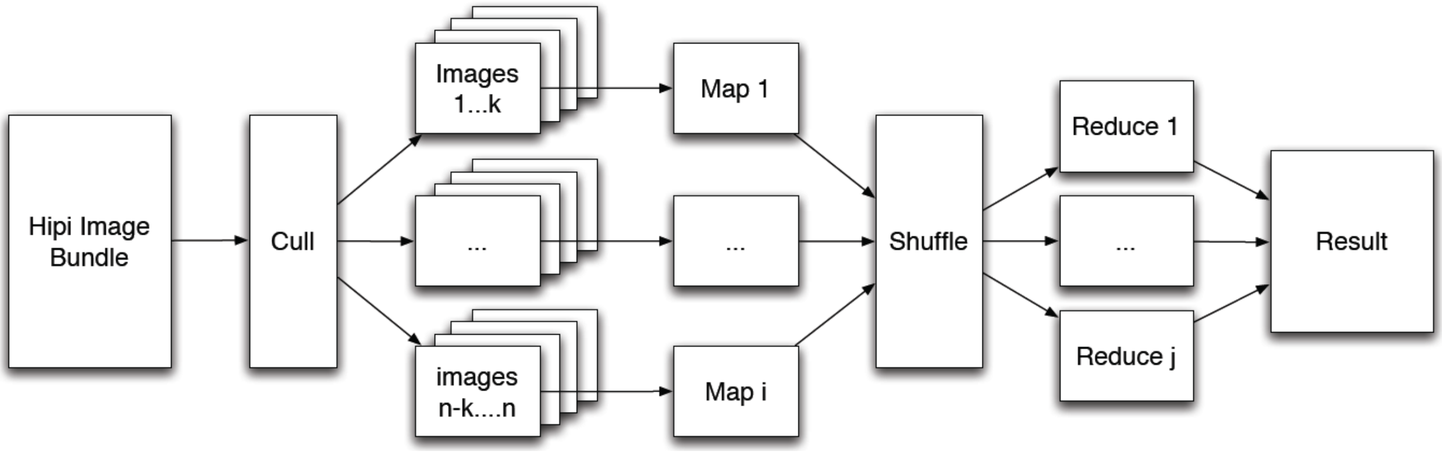


Fig 3: Hadoop Image processing Interface.

As a result, algorithms that evaluating the average value of all pixels in a collection of images can be implemented in few characters. Trimming is among the operations available for obtaining image patchwork. We had extracted such information from image pixels because it would be frequently desirable to access image headers and image files data while requiring to obtain image data. This is especially beneficial during in the filtering stage, including for applications which require metadata access, such as im2gis3. Through providing users with simple interfaces for obtaining access to the information required for image processing vision applications, MapReduce applications can be developed more effectively.

4.1 Experiment I

1) Implementation with Hadoop:

Portioning the images is the first job before taking them into the Hadoop's Map Reduce. Once splitting the image into memory-sized portions, a next step is to creating a MapReduce algorithm to operate the data. We had effectively established that each implementation of the method has essential data commonly accessible by determined overlaps or placing the parts inside the HDFS block size during the partitioning process. Moreover, because we can efficiently perform all required computations in the Map phase, we wouldn't need a 36 Reducer - Hadoop could be set to immediately transfer output to storage after the Map phase. As a consequence, the MapReduce programme in this situation comprises only of three concepts: InputFormat, Mapping, and OutputFormat. Their respective features are straightforward: read blocks from HDFS, transform them to Java objects containing the name, measurements, and number of pixels of the block's contents (each block contains one piece of the entire image), procedure these pieces with fast O(1) optimization technique, and finally transform the resulting objects back to PNG files and start writing to HDFS. Similarly to the previous section's practical example, the important factor is the picture's filename and the Values is a Java objects containing the image.

2) Testing

According to the Amazon Cloud official site, the ml.small and m2.xlarge instance categories do have the following parameters. An EC2 Computing Unit is about similar to a 1.0-1.2 GHz 2007 Opteron or Xeon Chipset.

Amazon EC2 cloud, all tests were run on a Hadoop cluster including one ml.small virtual environment as the master node and m2.xlarge virtual machines as compute nodes. In terms of Hadoop cluster configuration, numerous characteristics stayed at the default options in both runs with version 0.20.2 and 1.0.3. The two exceptions were limiting HDFS block size to 64 megabytes and reducing the maximum RAM for Map and Reduce jobs to 15 000 megabytes. The algorithm's requirement directly influenced the choice of m2.xlarge instances for computing nodes; all attempts to perform the experiment with 37 ml.small instance ended in failure due to a lack of storage.

Figure 1 demonstrates the main findings of the testing. To evaluate the method's MapReduce implementation towards its performance as a stand-alone ImageJ plug-in. To process all of the elements of the original picture consecutively in m2.xlarge example, we built a shell script which invoked an Image phrase. We can determine how much the Hadoop framework altered overall performance of the computations because the technical specifications of the instance were the same as those of the compute nodes. The decrease in performance, as seen in Figure 5. Considering that Hadoop also provides automatic failover, load balancing, and data distribution by itself, it may be maintained that this approach to image analysis has justified itself and could be used as a feasible option to similar problems.

Figure 5 represents the results of a performance analysis between Hadoop versions 0.20.2 and 1.0.3.

Table 1. Comparing performance between Hadoop version 1 and 2.

No of node	Wall time Version 1	Wall time Version 2
16	3030.01	3397
32	1668	1782

A comparison of response time between a Hadoop version 1 and 2 cluster. The latter instance provides an average of 5 testing sessions.

4.2 Experiment II

Figure 5 depicts experiment cluster configurations for both Hadoop and SGE. Each function requires the consumption with one CPU core and four gigabytes of RAM. A feasible experimental average bandwidth of 70 Mb/second, a disc reading rate of 100 Mb/second, and a write speed of 65 Mb/sec. Wall clock time and resource time are the measurements used to validate our proposed methods. SGE is being used experimentally as just a reference assessment.

Datasets

The experiment involves the use 5,153 T1 images acquired from healthy normal participants and obtained [19].

CASE I : Multidisciplinary unit

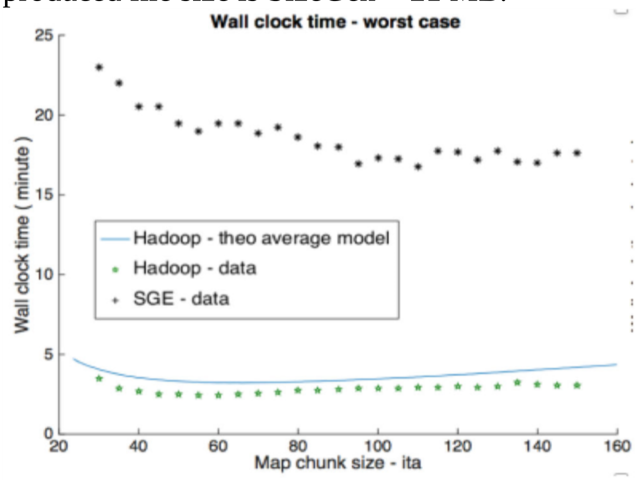
In [19], we observed that if data is transferred properly between embedded systems, Hadoop will consume more wall time then SGE. A new Hadoop Base-MIP load - balancing solution was defined by the amount and performance of CPUs per machine. Our objective is to experimentally demonstrate how well a network device can enhance Hadoop performance in a heterogeneous cluster. We implement the same experimental technique as [19], compression 5,153 T1 images to.gz format. The entire image input size is 77.4 GB, while the processing generates 45.7 GB of file format as output.

Regarding location of the data, each computer served as a Hadoop Data node and an HBase Regional Server [20].

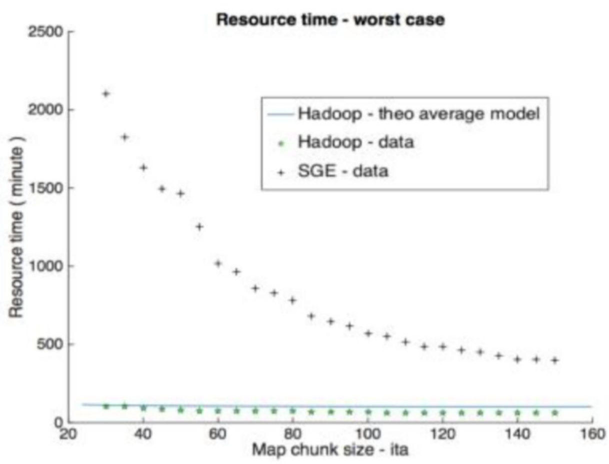
SGE has also been used to configure all devices. For both approaches machines acts as a cluster master.

CASE II - Analysing huge datasets

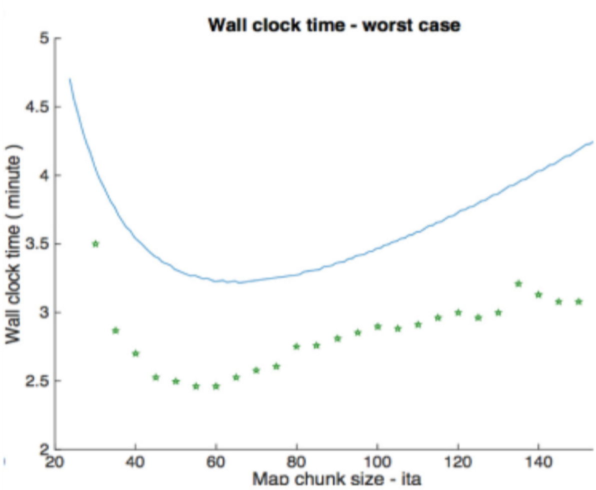
We initiated NiftyReg [21] to conduct rigorous affine transformations on all photos in order to register them to the MNI-305 space template [22, 23]. Our objective is to use the ANTS Average Images tool to average all 5,153 datasets. The produced file size is SizeGen = 21 MB. dataset's greatest file size is Size Big = 20 MB, the lowest file size is Small = 6 MB, and the average



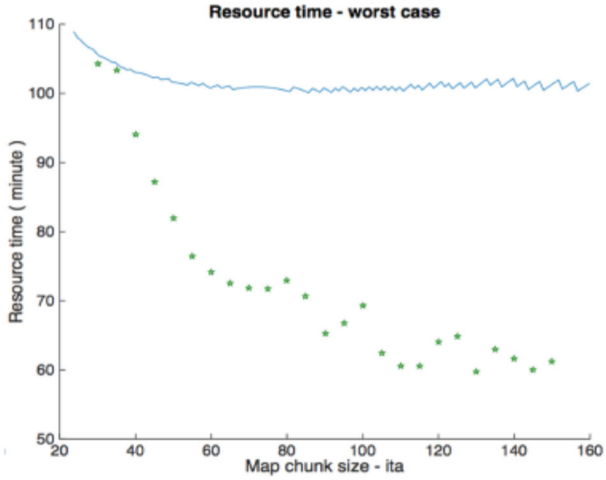
(A) Wall-clock time performance for Hadoop and SGE on large datasets analysis



(B) Resource time performance for Hadoop and SGE on large datasets analysis



(C) Wall-clock time performance for Hadoop and theoretical model



(D) Resource time performance for Hadoop and theoretical model

Fig 5: Results of a performance analysis between Hadoop versions 0.20.2 and 1.0.3.

CASE III – Rapid Nosql

Age and gender are two significant population-based investigation parameters that we are concerned on standardizing. Two techniques to Hadoop are developed and evaluated.

A simple approach is to keep data from all photos, indexes, and ages in the same column family. Overall performance of two Hadoop workloads is evaluated by the baseline SGE speed. For one mapping task, we established an empirical limit of 50 images per segment.

Conclusion

While cloud computing was used efficiently across product segments, there are still impediments to the use of Healthcare Image Processing. The far more significant challenge in the conversion to big data technologies is that the enormous volumes of information inside existing systems would not interact with others, as well as the data is available in various formats. A next difficulty for imaging data in the healthcare is maintaining patient confidentiality while keeping and exchanging information interconnected with appropriate connectivity. This is a difficult task for firms that create Predictive Analysis and Hadoop applications in keeping with National Indian Health Board regulations. It is certain that implementing the effectiveness of Big Data

Analytics and Hadoop's . Mapping Reduction into Medical Image Processing could increase the effectiveness of the algorithms, enabling researchers to provide more precise results in an efficient manner.

References

[1]. C.S. Lindquist, T.S.C. Lindquist, and T.V. Lindquist, "New image no processing algorithms requiring almost a priori design information," Signals, Systems & Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on, vol.2, pp. 1-4, November 1998.

[2]. D. Peter Augustine, "Leveraging Big Data Analytics and Hadoop in Developing India's Healthcare Services". International Journal of Computer Applications, vol. 89, no. 16, pp. 36-40. March 2014.

[3]. "Big data: 5 major advantages of Hadoop." [Online]. Available: http://www.itproportal.com/2013/12/20/big-data-5-major-advantages-of hadoop/ [Mar. 17, 2014].

[4]. In Kyu Park; Singhal, N.; Man Hee Lec; Sungdae Cho; Kim, C.W., #Design and Performance Evaluation of Image Processing Algorithms on GPUs." Parallel and Distributed Systems, IEEE Transactions on, vol. 22, pp. 91,104, January 2011.

[5]. Welch, E.; Patru, D.; Saber, E.; Bengtson, K., "A study of the use of SIMD instructions for two image processing algorithms," Image Processing Workshop (WNYIPW), 2012 Western New York, vol., pp.21, 24, November 2012.

[6]. Okuhata, H.; Imai, R. Ise, M.; Omaki, R.Y; Nakamura, H.; Hara, S.; Shirakawa, I, "Implementation of dynamic-range enhancement and super-resolution algorithms for medical image processing." Consumer Electronics (ICCE), 2014 IEEE International Conference on, vol.pp. 10-13 January, 2014.

[7]. U. Catalyurek, S. Hastings, K. Huang, V. S. Kumar, T. Kurc, S. Langella, S. Narayanan, S. Oster, T. Pan, B. Rutt, X. Zhang, and J. Sultz. Supporting large scale medical and scientific datasets. In Parallel Computing: Current & Future Issues of High-End Computing, Proceedings of the International Conference ParCo 2005, pages 3-14. 2005.

[8]. B. White, T. Yeh, J. Lin, and L. Davis. Web-scale computer vision using MapReduce for multimedia data mining. In Proceedings of the Tenth International Workshop on Multimedia Data Mining, July 2010.

[9]. Xu Zhengqiao; Zhao Dewei, "Research on Clustering Algorithm for Massive Data Based on Hadoop Platform," Computer Science & Service System (CSSS), 2012 International Conference on, vol., no., pp.43,45. 11-13 Aug. 2012

[10]. Hao, Chen; Ying, Qiao, "Research of Cloud Computing Based on the Hadoop Platform," Computational and Information Sciences (ICCIS), 2011 International Conference on, vol., no, pp.181,184, 21-23 Oct. 2011

[11]. "Hadoop Study Reveals Usage Stats, Benefits, and Challenges." http://architects.dzone.com/ [Feb. 9, 2014].

[12]. White, T. Hadoop: the Definitive Guide (2nd Edition) [M]. O'Reilly Media, 2010.

[13]. Markonis, Dimitrios, Roger Schaer, Ivan Eggel, Henning Maller, and Adrien Depeursinge. "Using MapReduce for Large-Scale Medical Image Analysis", 2012 IEEE Second International Conference on Healthcare Informatics Imaging and Systems Biology, 2012.

[14]. Weiyi Shang. Zhen Ming Jiang. Hemmati, H. Adams, B. Hassan, A.E. Martin, P. 2013. Assisting developers of Big Data Analytics Applications when deploying on Hadoop clouds, Software Engineering (ICSE), 35th International Conference. hipi.cs.virginia.edu/2014/2014.

[15]. Mukherjee, A. Datta, J. Jorapur, R. Singhvi, R. Haloi, S. Akram, W.2012. Shared disk big data analytics with Apache Hadoop, High Performance Computing (HiPC), 19th International Conference.

[16]. Karl Potisepp, "Large-scale Image Processing Using MapReduce", Thesis, Tartu University, 2013.

[17]. Huo, Y., Aboud, K., Kang, H., et al. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing; 2016. Mapping lifetime brain volumetry with covariate-adjusted restricted cubic spline

regression from cross-sectional multi-site MRI; p. 81-88.

[18].Chang F, Dean J, Ghemawat S, et al. Bigtable: A distributed storage system for structured data. ACM Transactions on Computer Systems (TOCS). 2008; 26(2):4.

[19].Apache HBase Team. Apache hbase reference guide. 2016. http://hbase.apache.org/book.html

[20].Ourselin S, Roche A, Prima S, et al. Block matching: A general framework to improve robustness of rigid registration of medical images. MICCAI 2000. 1935:557–566.

[21].Evans, AC., Collins, DL., Mills, S., et al. Nuclear Science Symposium and Medical Imaging Conference. IEEE; 1993. 3D statistical neuroanatomical models from 305 MRI volumes; p. 1813-1817.

[22].Mazziotta J, Toga A, Evans A, et al. A probabilistic atlas and reference system for the human brain: International Consortium Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2001; 356(1412):1293–1322. [PubMed: 11545704]