ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 8 , August 2017

Exploring Big Data Analytics for Satellite Imagery Data Using Hadoop Technique

[1] Ch.Rajya Lakshmi, [2] Dr.K.RammohanRao, [3] Dr.R.RajeswaraRao
[1] Assistant Professor,BVRITN, [2] Senior Scientist NRSC,Hyderabad, [3] Professor & Head,Jntukucev

***Abstract*— Now a days Big Data has defined very large amount of data, it includes both structured and unstructured format. The structured data analyzing is very easy task but an unstructured data analyzing is very difficult that can be produced by an individuals (eg. Twitter data)it also gathered by sensors(eg. satellites, videos) which can range from giga bytes, tera bytes and peta bytes. Big Data entitles more and more data that can be analyzed through various analyzing techniques. If the right analytic method is applied to unstructured datasets we can easily analyze and classifying various patterns, But at the same time will consider efficiency and scale of Data. In the real world the major issue of Big Data is early warning predictions is the use of Satellite imagery and Radar Sensor data. In the Satellite imagery data could reach a million derived spatial objects such data querying ,managing and various image patterns classification is very difficult task. So a proper architecture should be proposed to gain knowledge about Big Data for analyzing various Satellite imagery patterns classifications with hadoop technology. In the proposed architecture differentiate various classification methods for various satellite imagery pattern classification methods and also proposing Google's Map reduce C4.5 Algorithm for effective classification to increase performance of patterns classification and increasingly large volume of Data sets to results both time efficiency and scalabilty. This research is carrying based on NASA Satellite data and Twitter data and also in weather forecasting.**

***Keywords*— c4.5 ,Satellite imagery, Hadoop, Google's Map Reduce**

INTRODUCTION

Big data analytics plays a very Important role in optimal storage of both structured and unstructured data storage. Normally remote sensing accumulates very large amount of data in the form of multispectral and high resolution satellites images. The big data analytics techniques classifies very large amount of data, generally big data generated by satellites, twitter and face book. All these data storage by using data base management system is very difficult and classification of imagery data from the satellite is not an easy task. Nowadays extracting useful data from the large amount of data is very interesting task. Big Data has been seen and defined and definition has been provided by Kitchin (2013) as Data Sets that are high in Volume Velocity and variety. When the right analytics are applied to data sets, then only important information can be captured and information can be captured and predict the future risks and to inform the disaster risk and management .

II. LITERATURE SERVEY

In the recently previous research they used the different classification algorithms to classify the satellite images by using classification techniques and machine learning paradigms can identifies the decision making and predictions but those predictions are not accurate .In recent trends one of the most popular application of big data is GIS,It handles complex and large amount of spatial data objects through satellite images[2].In the recent trends GIS is a challenge research topic due to large size of spatial images that includes various image patterns that can manage and storage is very challenge task[1].In the previous research they have classified images for flah flood images[5] by using that they found The affected population number and land area are determined and compared. In the previous reserch the subject of images, Yan et al. [9] proposed the algorithm of extensibility based on MapReduce model which has been tested on 260,000 images. Almeer previously designed and implemented a system for remote sensing image processing systems with the help of Hadoop and cloud computing systems with 112 compute nodes [10].

III.PROPOSED RESEARCH

In the proposed method we propose a new architecture that is Hadoop Mapreduce technique with c4.5 classification algorithm.By using this technique can process very complex imagery data with good classification algorithm after this classification we can compare the current result with previous result we can get scalable and cost effective and efficient results. The most challenging task is handling of Remote sensing data that data is generated by the aircrafts,satellites and any other sensing device all these data collecting and managing is a Big Data problem. Because this data includes high resolution pictures and having various patterns and various standard data formats[3] all these

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 8, August 2017

classification is very difficult that comes under "Big Data". The total data handling and computing need a very good efficient and parallel computing architecture that is googles map reduce Hadoop technique with efficient c4.5 classification technique for handling of satellite imagery data. In Machine learning during the test phase selected regions are tested with different classifiers[8].

IV. METHODOLOGY

This Research is conducting with Hadoop and Goggles Mapreduce with c4.5 algorithm. Hadoop is a parallel distributed frame work. Initially Collected Data includes noisy and various hidden patterns are included all these can be handled by using Hadoop map reduce with c4.5 .Normally Satellite image classification can be done in supervised learning and un supervised learning methods by using supervised methods data already classified with k-means and LDA [13] techniques. This satellite images using Kmeans Algorithm[5] .In the proposed research one architecture proposed with Various levels those are given below

LEVEL1:Handle Big Data Set(it includes both structured and unstructured data)

LEVEL2:Data Filtering(Separate both structured and unstructured data)

LEVEL3:Data Analysis and Data Distribution through Hadoop Frame work.

LEVEL 4:Data Processing in two different data sets.

LEVEL 5:Classify Previous results with present architecture results.

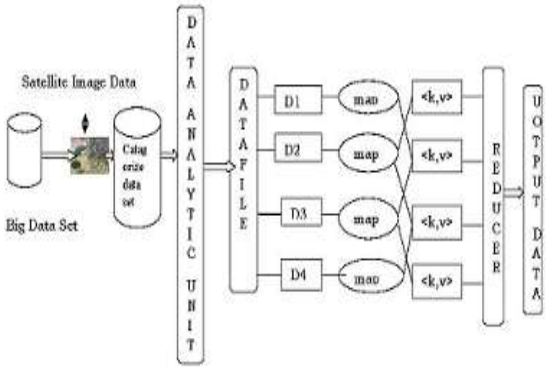

Fig. IV.I Proposed Architecture with Satellite Image

To process a large amount of satellite imagery data efficiently this data is fed to Hadoop distributed file system. The resultant signature image as as shown in Fig IV.I. It is necessary to divide these higher resolution images into multiple segments and assign each image segment to different slave machines to efficiently compare the images. This can be done in distributed environment. Hadoop is a distributed frame work model after google mapreduce to process large amount of data in parallel[3].For Data classification we can take any type of satellite images. By using t images we can find out weather forecasting and floods impact to the nature and also find out future predictions. In the previous research Satellite images processed by using various techniques but proposed technique will give better than the previous results.

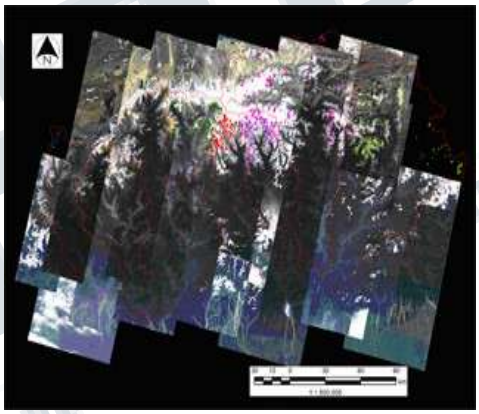

Fig IV.II Segmented image[4] with HDFS

V .CLASSIFYING IMAGERY DATA WITH C4.5 AND PRUNING

C4.5 Algorithm is classification model. In this at each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. In the Hadoop google's map reducing technique the entire data parallel distributed among all the nodes after that the resultant Data can be classified by using c4.5 algorithm. After classification data can be pruned for the reduction of the classification errors.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 8, August 2017

Hadoop Apache Log Processing with Cascading values			Training Data Set for C4.5 decision tree algorithm	Result with pruning
1 Node	Runtime Sec/MB Sec/MB/Node	21m46s 0.127 0.127		
4 Nodes	Runtime Sec/MB Sec/MB/Node	8m3s 0.0471 0.0157		
20 Nodes	Runtime Sec/MB Sec/MB/Node	1m30s 0.00878 0.000585		
Naive Perl	Runtime Sec/MB Sec/MB/Node	41m48s 0.251 0.251		

Fig V.I Hadoop Apache Log Processing Cascading values The Main purpose of this technique is implement a scalable modeling system through Hadoop MapReduce framework is used for training the classifiers and performing subsequent image classification with c4.5 with pruning. The proposed system will follows the given Algorithms for the scalable, future extraction and cost effective imagery classification systems.

V.II Algorithm for satellite imagery data classification

Step 1: Take Satellite database in the form of Big Dataset

Step2: Extract Satellite Images from the Database

Step3:Categorize Data Set

Step4:Apply Data set to Analytical Unit through Hadoop frame work.

Step5: On HDFS, we execute set of operations parallel by using

MapReduce Programs

Step6:Apply c4.5 decision tree with pruning on the mapreduce programs

Step7.Compare present result with previous results

V.III Algorithm for C4.5 for Imagery data classification

Input:Training Dataset P; attributes R
Output:Decision Tree
Step1:IF P is NULL Then
Step2:retrn failure
Step3:end if
Step4:set Tree={ }
Step5:for a € to R do
Step6:set Info(a,P)=0,and spilt info(a,P)=0
Step7:Compute Entropy(a)
Step8:for v € to values(a,P) do
Step 9:Compute Info(a,p)
Step 10:Compute splitInfo(a,P)
Step11:end for
Step 12:Caluculate Gain(a,P)
Step13:Caluculate Gain Ratio(a,P)
Step14:end for
Step15:Find a(best) and attach to Tree
Step16:for v€ values(a(best),P)do
Step17:call c4.5(T(a),v)
Step18:end for
Step19:Return Tree

By applying all these algorithms we will get scalable and cost effective results in an efficient manner. Generally C4.5 algorithm uses information gain as splitting criteria[10] . It can accept data with categorical or numerical values. To handle continuous values it generates threshold and then divides attributes with values above the threshold and values equal to or below the threshold. C4.5algorithm can easily handle missing values by using pruning.

VI.CONCLUSION AND FUTURE RESEARCH WORK

In this research paper, we explore the big data analytic tools integrating with Hadoop mapreduce technique for satellite imagery because the utility and potentiality of big data for satellite imagery data is growing because those data sets includes various heterogeneous image patterns are included. by using Big Data analytic mining algorithms in image classification which is analyzed based on higher Image classification and improve system performance through Hadoop map reduce with best c4.5 algorithm . In future we planned to apply this work with various Datasets like Twitter , facebook and various social networks and this work is also carried out on NASA Satellite Data Sets and also used in weather forecasting.

VI. REFERENCES

[1] .I.Chebbi,w.boulila,I.R.Farah "Improvement of satellite image classification approach based on Hadoop/Map Reduce" in IJESRT ,Volume 3 , January 2017,pp 435-442.

[2] .Ablimit Aji, Fusheng Wang, Hoang Vo, Rubao Lee, Qiaoling Liu, Xiaodong Zhang, andJoel Saltz "Hadoop-GIS: A High Performance Spatial Data WarehousingSystem over MapReduce"Proceedings VLDB Endowment. Author manuscript; available in PMC 2013 October 31.

[3] .Yaman Kumar Gunturi*, K. Kishore Raju "RealBDA: A Real Time Big Data Analytics For Remote Sensing Data By Using Mapreduce Paradigam" in IJESRT ,Volume 3 , January 2017,pp. 435-442.

[4] .Sarade ShrikantD.,Ghule Nilkanth B.,Disale Swapnil
P sane Sandip R. "Large Scale Satellite Image Processing Using Hadoop Distributed System" in IJARCET Volume

3 Issue 3, March 2014 pp. 731-735.

[5] . Ismail Elkhrachy "Flash Flood Hazard Mapping Using Satellite Images and GIS Tools: A case study of Najran City, Kingdom of Saudi Arabia (KSA)" in The Egyptian Journal of Remote Sensing and Space Sciences 4 August 2015 pp 261–278

[6]. Vaibhavi S. Shukla, Jay Vala in "A Survey on Image Mining, its Techniques and Application"in IJCATM Volume 133 – No.9, January 2016,PP.12-15.

[7] P. Chandarana and M. Vijayalakshmi, "Big Data analytics frameworks," in Proc. Int. Conf. Circuits Syst.Commun. Inf. Technol. Appl. (CSCITA), 2014,pp 430–434.

[8]. PThamilselvan, Dr. J. G. R. Sathiaseelan "A Comparative Study of Data Mining Algorithms for Image Classification" in I.J. Education and Management Engineering, 2015, 2, 1-9 DOI: 10.5815/ijeme.2015.02.01

[9] Yan, R., Fleury, M.-O., Merler, M., Natsev, A. and Smith, J.R. (2009) Large-Scale Multimedia Semantic ConceptModeling Using Robust Subspace Bagging and MapReduce. Proceedings of the 1st ACM Workshop on Large-ScaleMultimedia Retrieval and Mining, Beijing, 23 October 2009, 35-42.

[10] Almeer, M.H. (2012) Cloud Hadoop MapReduce For Remote Sensing Image Analysis. Journal of Emerging Trends in Computing and Information Sciences, 3, 637-644.

[11] Jiong Xie, Shu Yin, Xiaojun Ruan, Zhiyang Ding, Yun Tian,James Majors, Adam Manzanares, and Xiao Qin "Improving MapReduce Performance through Data Placement inHeterogeneous Hadoop Clusters" in Proc. 19th Int'l Heterogeneity in Computing Workshop, Atlanta, Georgia, April 2010.

[12] Cohen, S. E. 2013. Sandy Marked a Shift for Social Media Use in Disasters. Emergency Management http://www.emergencymgmt.com/disaster/Sandy-Social-MediaUs in Disasters.html#.UTkc9hoaKdo.facebook (last accessed 9 March 2013).

[13]. Beth Tellman (1), Bessie Schwarz (2), Ryan Burns (3), Christiaan Adams (4) "Big Data in the Disaster Cycle: Overview of use of big data and satellite imaging in monitoring risk and impact of disasters" in UN Development Report 2015.

[14] Joyce,K.E., Belliss,S.E., Samsonov,S.V., McNeill,S.J., Glassey,P.J.(2009)".A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters" in Progress in Physical Geography,33(2), 183207.doi: 10.1177/0309133309339563

[15] Er.Navjot Kaur, Er. Yadwinder Kaur "Object classification Techniques using Machine Learning Model In (IJCTT) – Volume 18 Number 4 – Dec 2014,pp 170-174.

[16]. Anil K Goswami, Swati Sharma, Praveen Kumar "Nearest Clustering Algorithm for Satellite ImageClassification in Remote Sensing Applications" in IJCSIT Vol. 5 (3) , 2014, 3768-3772,pp 3768-3772.