International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 06, June 2016
www.ijates.com
ijates
ISSN 2348 - 7550

REAL-TIME BIG DATA ANALYTICAL ARCHITECTURE FOR IMAGE PROCESSING FRAMEWORK

Smita Patil[1], Dr. Rekha Patil[2]

[1,2]CSE, PDA College of Engineering Kalaburgi/ VTU, (India)

ABSTRACT

Big Data analytics and Hadoop architecture has significantly improved the real time data processing and storage Data with significant amount of velocity, veracity and variety was in the earlier processed by Data warehouse. The problem in the Data warehouse technique is that it always used to construct a sample out of the entire Data so, any given instant of time only part of the Data has used to get analyzed at any given instance. However, with the introduction of HDFS and Hadoop file system. We are now capable of analyzing and storing large amount of Data in near real time latency response. Many past work has proposed gathering, analysis and interpretation of data from varies different sources includes Webs, e-commerce Data, physical sensors Data and many more. Even though, certain image processing algorithm is been proposed over Big Data architecture in the past most significant work proposes an overall image processing frame work in conjunction with Big Data with either similar of dissimilar properties can be analyzed through Big Data architecture in this work we have proposed a novel real time approach for extracting the features and performing processing operation on large volume of images in a distributed Big Data frame work. The proposed system distributes large volume if images by splitting them into smaller parts through a broker node into multiple physical Data nodes process the chunk of the images in parallel at individual node and extract their features. This feature is stored at this data node as well as sends back to the broker (name node). The name node keep track of the entire split a through a metadata system and data record result shows that improves he image processing speed and efficiency significantly over serial execution of either single or multiple node processing the data either serially or in parallel way.

Keywords: Big Data, HDFS, distribution.

I. INTRODUCTION

Traditionally the data is managed, integrated in databases whose storage capacities were in terms of megabytes and gigabytes. The traditional databases can store and manipulate only structured data containing rows and columns. Now days the data is rapidly growing in terms of terabytes and further in pet bytes. Traditional databases cannot handle such huge datasets. To overcome this major drawback, BIGDATA has come into existence. Data is row records. Facts and statics collected together for reference or analysis. Which may or may not convey the any message and meaning extracted from the row data is information by using certain algorithms so, data management is an essentially record management however information management is essentially

method applied on data for extraction of the meaning? Data is row records. Facts and statics collected together for reference or analysis. Which may or may not convey the any message and meaning extracted from the row data is information .but by using certain algorithms so the data management is an essentially record management however information management is essentially method applied on data for extraction of the meaning?

Big data is refers to datasets whose size are beyond the ability of typical database tools to capture, store, manage and analyses. So anything in its own category more than the average we call it as a big. System that allows storing hedge amount of ever changing records and analysis them in a real time, as the records are changing the analysis is taking care of the same change at the same time at every input output of the record.

"Big Data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store manage and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to subjective big data i.e., we don't define big data terms of being larger than a certain number of terabytes. We assume that, as technology advances overtime can vary by sector, depending on what kinds of software tools are commonly available and size of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple pet bytes. The ability to store, aggregate and combine data and then use the results to perform deep analyses has become ever more accessible as Moore's Law in computing, its equivalent in digital storage and cloud computing to lower cost and other technology barriers. Big data is any attribute size being one of them, that challenges constrains of a system capabilities or a business need

Recently ,a great deal of interest in the field of Big Data and its analysis has risen[10], mainly driven from extensive number of research challenges strappingly related to bonafide applications, such as modeling, processing, querying, mining, and distributing large-scale repositories. The term "Big Data" classifies specific kinds of data sets comprising form- less data, which dwell in data layer of technical computing application and the Web[11]. The data stored in the underlying layer of all these technical computing application scenarios have some precise individualities in common, such as a) large- scale data, which refers to the size and the data warehouse; b) scalability issues, which refer to the application's likely to be running on large scale(e.g., Big Data) c) sustain extraction transformation loading (ETL) method from low, raw data to well thought-out data up to certain extent and d) development of uncomplicated interpretable analytical over Big Data warehouses with a view to deliver an intelligent and momentous knowledge for them. Big Data are usually generated by online transaction, video/audio, email, number of clicks, logs, posts, social network data, scientific data, remote access sensory data, mobile phones, and their applications [12]. These data area accumulated in data base that grow extra ordinarily and become complicated to confine, form, and store, manage, share, process, analyze, and visualize via typical database software tools.

Advancement in Big Data sensing and computer technology revolutionizes the way remote data collected, processed, analyzed, and managed. Particularly, most recently designed sensors used in the earth and planetary observatory system are generating continuous stream of data. Moreover, majority of work have been done in the various fields of remote sensory satellite image data, such as change detection, gradient-based edge detection, region similarity- based edge detection, and intensity gradient technique for efficient intra prediction. High-speed continuous stream of data or high volume offline data to "Big Data," which is leading us to a new world of challenges. Such consequences of transformation of remotely sensed data to the scientific understanding are a critical task. Hence the rate at which volume of the remote access data is increasing, a number of individual

users as well as organizations are now demanding an efficient mechanism to collect, process, and analyze, and store these data and its resources.

The rest of the paper is organized as follows. Section II presents the related work. Section III gives out details of the proposed scheme. The implementation and results are discussed in section IV, then finally concludes the paper and outlines future works in section V.

II. RELATED WORK

Design and implement satellite remote sensing integrated application service platform, [1] proposed platform has tried to alter the traditional remote sensing application service mode by means of combing it with E-commerce so that it can offer data services, product services and online analytical services. Analyzing Utilization Rates in Data Centers for Optimizing Energy Management, [2] Energy data analysis with the proposed equations for performance measurement and forecasting, corroborated by evaluation with real data in a university setting. Big Data Analytics in the Public Sector, [3] the framework proposed for the development of new decision support systems that integrate "past data" with "real time data". The methodology emphasizes induction with the collection of qualitative data in order to move from observed facts to theory. Remote Sensing Processing: From Multicore to GPU,[4] The proposed framework that has proven to be efficient with standard implementations of image processing algorithms and it is demonstrated that it also enables a rapid development of GPU adaptations. A Survey of Clustering Techniques for Big Data Analysis,[5] Clustering is one of the major techniques used for data mining in which mining is performed by finding out clusters having similar group of data. Comprehensive analysis of these techniques is carried out and appropriate clustering algorithm is provided. A Workflow Model for Adaptive Analytics on Big Data. On Traffic-Aware Partition and Aggregation in Map Reduce for Big Data Applications, [6] a decomposition-based distributed algorithm is proposed to deal with the large-scale optimization problem for big data application and an online algorithm is also designed to adjust data partition and aggregation in a dynamic manner. Analyzing Utilization Rates in Data Centers for Optimizing Energy Management, [7] proposed equations for performance measurement and forecasting, corroborated by evaluation with real data in a university setting. An adaptive framework for the execution of data intensive Map Reduce applications in the Cloud. A Proposed Secure Framework for Safe Data Transmission in Private Cloud, [8] helps in protecting the data from unauthorized entries into the server, the data is secured in server, based on users' choice of security method; so that data is given high secure priority without affecting the lower layer. intensity gradient technique for efficient intra prediction[9], the high speed continuous stream of data or high volume offline data to "Big Data," which is leading us to a new world of challenges

III. PROPOSED WORK

A real time big data analytical architecture we assume that we have a large volume of satellite images coming to us now this images are to be processed it could be anything it could transformation applied on the images or extracting the images. Conventionally speaking what could we could have done is we would have taken a single images would have performed the transform. And imaging itself combination of pixels so larger the image it going to take that much of time to process it then we are going to set the result and instead of doing that we are Going to use big data image processing framework. Will have a name node so ideally every communication on

HDFS is carried out on the top of the any network so data is encrypted and HDFS provides the security over the traditional data base. Here objective of the work is to process hedge amount of data using multiple nodes. Using big data architecture reduces the overall processing and analyzing time required for images processing. There for our work is to provide efficient robust and low latency solution for processing hedge volume of data.

Our proposed frame work has a name node and data node firstly name node it takes a one image into the account and it will know the available data node then name node divide the image and sends it to the available data nodes. Those data node process it and store it in their local memory itself. Name node will be having metadata in their memories which contains the information which data node is having which part of the image. If any user what the result he can directly contact to the data nodes and get the result.

We create a local Wi-Fi network by configuring n number of personal computers. Out of the one computer one is a name node others is data node. The system which runs a name node can also be run as data node. Job n task pattern, we directly use images as job the objective of system is to process the image in a distributed way we consider gray scale of the image and further its histogram statistics be obtained in statistical or analytical major of the images. protocol we use multicasting as under need protocol in order to meeting at the data the broker into multiple nodes HDFS file system as well as flat file metadata services at the name node.

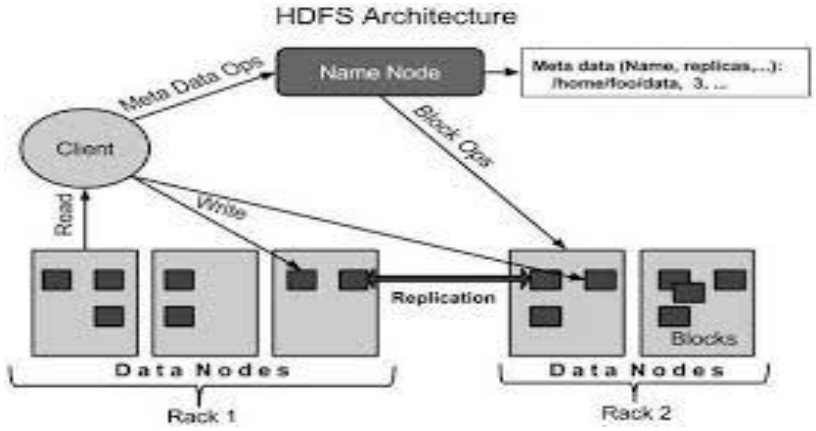

Fig.1 System Architecture

Detailed methodology firstly we network is created with data node and name node. the next all the IP address of data node enter into a name node for the virtual connection and then name node creates the e metadata node which are connect with and directly images are given to name node it distributes the images into mutli part then transmits each data node convert the image into gray scale and creates histogram and stores. once it get the response then metadata updates the storing info about the particular file in order to demonstrate of the entire process we process by collecting back images at the name node this mages stored the gray scale converted images at name node .

IV. IMPLEMENTATION AND RESULT ANAYSIS

Implementation of this work is done by using tool called MAVEN repository. A repository in MAVEN is used to hold build artifacts and dependency of varying types. And we have used Hadoop core that is 1.2.1version big data hold the key value to understand the analytics of the data.

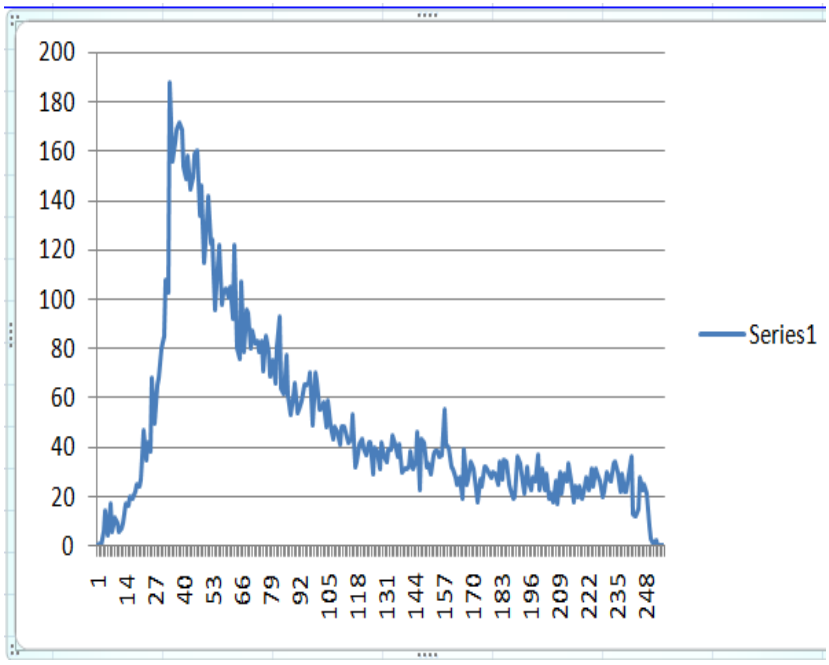

Figure.2 Analysis of the single image

As shown in above graph which represents the analytical presentations of the single image with single data node. And the statics of the processed values are have been processed and save over the each data node and result is saved in the sever node. Graph is represented using low intensity to high intensity nothing but a gray scale image will be having 256 color values that is 0 to 256. Histogram is graphs that tell you how many number of color vales are appeared.

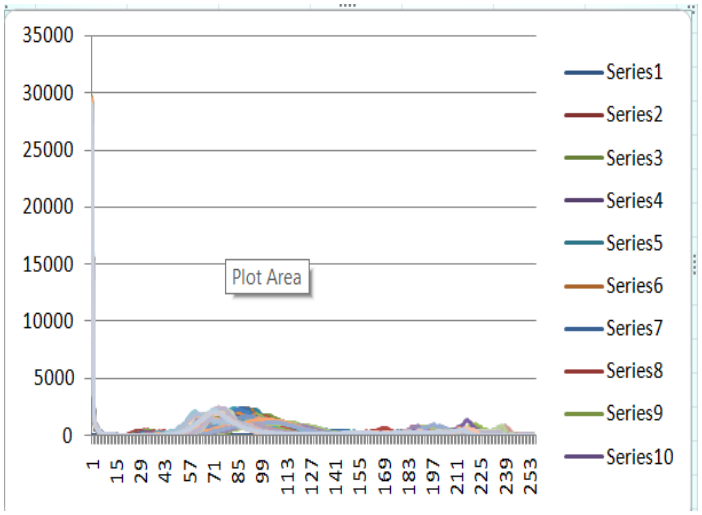

Figure.3 Analysis of the large set of images

As shown in above figure.2 and figure.3 x axis represents the color values and y axis which represents name of the image color submitting is changed so entire directory or image histogram is represented in the graph. This is a real time data combine together at the server end. Series one represents the values of first color and second color all the different images so on. This is done by using four data node and hedge images.

It provides high scalability. As the amount of data increases any number of computer are added into framework without required special configuration at protocol level. Robustness by default it keep three copies of same data by that data loss is reduced. Result show over method is extremely efficient and reduces the latency in comparison to single node base system.

V. CONCLUSION

With the increasing number of satellite images a large volume of created a new research area several of the agencies that the processing satellite images has been traditionally adapting greedy computing architecture. Where number of node are configured to perform a single job. However, one the drawback of such a system is that even though processing is disturbing and collaboration is integrated into single node. However distributed system improves their greedy framework by extending the collaboration and query service also into distributed node. In such system when a name node or data node receive the request for certain data it just distribute the into those node which hold the data into name node can directly dissipate the data into client requesting for the data .In this work we have introduced an images processing services for satellite images as the histogram analysis of this images. Which is plays an important role we have offered a gray scale conversion and histogram analysis. Result show over method is extremely efficient and reduces the latency in comparison to single node base system.

Our work can be further improved by incorporating to other image processing and analytical services into the existing framework and the then extending the framework with machine learning techniques that components like neural network support network technique trained or tested with data packet.

REFERENCES

[1] Yang Banghui and Chi Tianhe," A Preliminary Study on the Design and Implementation of Satellite Remote Sensing Integrated Application Service Platform" 2009 IEEE

[2] Michael Pawlish and Aparna S. Varde," Analyzing Utilization Rates in Data Centers for Optimizing Energy Management" lEE International Green Computing Conference 2012

[3]Joni A. Amorim, Sten F. Andler, Dr and Per M. Gustavsson, Dr Big Data Analytics in the Public Sector', 2013

[4] Emmanuel Christophe, "Remote Sensing Processing: From Multicore to GPU" VOL. 4, NO. 3, SEPTEMBER 2011 IEEE

[5]" Saurabh Arora and Inderveer Chana," A Survey of Clustering Techniques for Big Data Analysis" 2014 IEEE

[6] Huan Ke ,Peng Li and Song Guo," On Traffic-Aware Partition and Aggregation in MapReduce for Big Data Applications" 2015 IEEE

[7] Michael Pawlish and Aparna S. Varde," Analyzing Utilization Rates in Data Centers for Optimizing Energy Management" lEE International Green Computing Conference 2012

[8] Rohit Maheshwari, Sunil Pathak "A Proposed Secure Framework for Safe Data Transmission in Private Cloud" , Volume-1, Issue-1, April 2012 IEEE

[9]A.-C. Tsai, A. Paul, J.-C. Wang, and J.-F. Wang, "Intensity gradient technique for efficient intra prediction in H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 5, pp. 694–698, May 2008.

[10] D. Agrawal, S. Das, and A. E. Abbadi, "Big Data and cloud computing: Current state and future opportunities," in Proc. Int. Conf. Extending Database Technol. (EDBT), 2011, pp. 530–533.

[11] K. Michael and K. W. Miller, "Big Data: New opportunities and new challenges [guest editors' introduction]," IEEE Comput., vol. 46, no. 6, pp. 22–24, Jun. 2013.

[12] C. Eaton, D. Deroos, T. Deutsch, G. Lapis, and P. C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. New York, NY, USA: Mc Graw-Hill, 2012