Big Data Architecture for Remote Sensing Applications

Arockia Panimalar.S[1], Subhashri.K[2]

1,2 Assistant Professor, Department of BCA & M.Sc SS, Sri Krishna Arts and Science College, Tamilnadu

---***---

Abstract - *Big Data is the new experience bend in the new economy driven by enormous data with larger volume, speed and assortment. The real time application preparing of remote sensing of expansive volumes of Big Data appears at to begin with, and provide extracting and analysis in valuable information with an effective manner drives a propelled system toward a novel computational difficulties, for example, to examine, aggregate and store, where data are remotely accumulated. The proposed architecture can store raw data for the analysis of the offline data when required. It comprises of three main units, such as 1) Remote Sensing Big Data Acquisition Unit (RSDU), 2) Data Processing Unit (DPU) and 3) Data Analysis Decision Unit (DADU). First, RSDU acquires data from the satellite and sends this data to the Base Station, where initial processing takes place. Second, DPU assumes a fundamental part in architecture for proficient processing of real-time Big Data by giving filtration, load balancing, and parallel processing. Third, DADU is the upper layer unit of the proposed structure, which is obligated for compilation, storage of the outcomes, and generation of determination founded on the outcome obtained from DPU. The proposed architecture has the capacity of isolating, load balancing, and concurrent processing of just valuable data. Consequently, it result in adequately analyzing real time remote sensing tremendous data utilizing earth observatory method. Also, the proposed big data architecture gives the limit of putting away and making analysis of incoming raw knowledge to perform offline evaluation on mostly stored dumps, when required. Eventually, a targeted analysis of remotely sensed earth observatory tremendous data for land and sea field is provided making utilization of Hadoop. What's more, more than a couple of algorithms are proposed for every single phase of RSDU, DPU and DADU to identify land and sea.*

*Key Words***: Data Analysis Decision Unit (DADU), Data Processing Unit (DPU), Hadoop, MapReduce, Offline, Realtime, Remote Senses, Remote Sensing Big Data Acquisition Unit (RSDU).**

1. INTRODUCTION

Advancement in Big Data sensing and computer technology revolutionizes the way remote data collected, processed, analyzed, and managed in effective manner [1]. Big Data are customarily generated by way of online transaction, video/audio, electronic mail, number of clicks, logs, posts, social community information, scientific knowledge, far away entry sensory information, cellular phones, and their purposes [6][7]. These data are collected in databases that grow extraordinarily and end up complex to confine, type, retailer, control, share, process, analyze, and visualize via traditional database software tools. Such process incorporates powerful investigation of the data itself and ability of tools used to analyze it. Accordingly, to the technology capabilities, tens hundreds of terabytes storage are needed to handle big data. Therefore, big data and related analysis are very essential in modern science and business [2].

The advanced technology in the big data gives a way to the remote data, which can be collection, managing, analyzing and processing. Recently designed remote sensors that are used for the earth observatory streams the data continuously and generates large amount of data. Most of the works have been accomplished in the special fields of far flung sensing data from the satellite TV for PC, similar to gradient centered part detection [4] and change detection [5]. This paper is concentrated on the high speed continuous real time streaming data or large amount of offline data i.e. big data, this leads to a new challenge. Such consequences for scientific understanding of transformation of the remote sensed data are critical task [6] [3]. Certainly, most lately designed sensors employ in the earth and planetary observatory method are producing continuous movement of knowledge. In addition, majority of labour had been executed in the various fields of far flung sensory satellite TV for PC photo knowledge, similar to change detection [8], gradient established part detection [9], area similarity established side detection [10] and intensity gradient manner for effective intra prediction [11].

The splendid growth in the data additionally posing new challenges, such as aggregate gigantic quantity of data? How user can store such data in a limited quantity of reminiscence allocated for the special challenge? Additionally, the best way to procedure and analyze these data when there is no such smart algorithm is available? In addition, significant-scale data can't be tackled through standard discount strategies on the grounds that their runtime becomes impractical [12]. Having giant-volume knowledge, all of this has to occur in a mechanized method considering it requires various information constitution as well as semantics to be articulated in various forms of computer readable data formats.

However, by analyzing simple data having one data set, a mechanism is required of how to design a database. There might be alternative ways to store all of the same information. In remote access networks, the place the data

starting place such as sensors can generate an overwhelming quantity of raw information i.e. information acquisition, in which much of the information are of no meaning that can be filtered or compressed by way of orders of magnitude. In order to use such filters, they don't discard priceless know-how. Generally, the data gathered from remote areas are not in a format ready for analysis. Therefore, the 2nd step does knowledge extraction, which pulls out the significant information from the underlying sources and switches it in a structured formation suitable for analysis. For instance, the data set is covert to a single-category label to facilitate analysis, despite the fact that the very first thing that we used to believe about massive data as constantly describing the fact. However, this is far away from reality. Sometimes we have to deal with erroneous data too, or some of the data might be not clear.

2. RELATED WORK

This segment provides the detail summing up of the previous work done in the remote sensing real time big data. The digital world generating the high amount of the data continuously, current technology and the tools to store and analyze the large amount of processed data not an easy task, since the data is not able to extract the needed data sets. So there is a need of an architecture that can analyze both the offline data as well as real time data sets. There is a persuasive advantage in the business undertaking by getting the required information from the big data than ample data sets.

Understanding the earth atmosphere or condition requires extensive volume of data or information accumulated from various sources, for example, air and water quality monitoring sensors, amount of oxygen, CO2 and the other gases present in the air, remote access satellite for observing the characteristics of the earth et cetera. In the healthcare scenarios, there is vast amount of the information about the medicines, patients, medical history and different points of interest accumulated by the medical specialist. The previously mentioned data is exceptionally perplexing in nature, there is an odds of missing the vital data. Day by day the data becoming very large by social networking, online streaming, system logs, mails and remote data, it will be exceptionally hard to process massive amount of data. Fundamental issue is the means by which to store the extensive amount of data i.e. big data and what data is to keep and what data is to be discarded, extricating the valuable data from the big data is the challenging task [2].

The vast majority of the data is created by the streaming data. In data stream model, the data will arrive at a very high speed and the algorithm needs to process them. This data stream causes several challenges in designing the data mining algorithms. First, algorithm needs to make utilization of less number of resources. Second, it can manage data that can change after sometime. Resources are overseen in a

proficient and minimal effort path by the green computing [7].

Green Computing is detailed process of study to use the available computing resources in an efficient way depending upon requirement. Here, the problem is not only the scaling issue but also error handing, lack of structure, heterogeneity, privacy, visualization and timeliness. The test is to plan a high performance computing systems that can be capable incorporate resources from various areas. Despite the fact that the cloud computing systems demonstrated a high level performance for Remote Sensing (RS) applications, there are challenges as yet remaining with respect to energy and the time utilization. The big challenge emerges when collecting and the managing RS big data. The RS data are collected from spacecraft, airplanes, satellite and other sensing devices.

Remote sensing data growing explosively, we have entered in the period of very high resolution, observation of the earth. Remote sensing data also considered as a "Big Data". With the propel sensors we can take even high spatial resolution images, spectral resolution and furthermore temporal resolution. The headway in the technology of the PCs and the remote sensing devices expands an enormous development sensing data [9]. Streamed from the spacecraft was approximately around 1.7GB, this data is gathered by a single satellite and increased numerous terabytes every day. The worldwide records of observatory data of the earth would surpass to one Exabyte as indicated by the OGC statistics.

Different standard format data sets of remote sensing are put away in structured files, the formats including ASCII, HDF, netCDF et cetera. Diverse organization have distinctive standard format of the data sets, distinctive format of data has its own format libraries and operation interfaces. Huge amount of data need to compute in an efficient way and only the useful information need to be extracted from the big data. So there was a need of the architecture for filtering the data, load balancing, aggregating and the decision analysis.

3. METHODOLOGY

The notion of authorized data deduplication was proposed to defend the data security by together with disparity privileges of users in the duplicate check. We also presented various new deduplication constructions behind authorized replacement check in hybrid cloud architecture, in which the duplicate check tokens of files are generated by the private cloud server with private keys. Security analysis showcase that our schemes are at ease in phrases of insider and outsider attacks targeted within the proposed safety mannequin. We applied a prototype of our projected authorized duplicate verify scheme and carry out experiments on our prototype. We demonstrate that our authorized duplicate check scheme incurs smallest below evaluate to convergent encryption and network move.

4. PROPOSED WORK

Even the big data has the variety technologies as in cloud computing. System architecture is considered as the core of any proposed system fetched from analysis and design. The architectural diagram for our system is shown in Fig 1 and it contains the following components. Generally the server acquires the data from the satellite and thus provides a backup at servers connected directly to satellite. These servers are unauthorized and highly preserved from public accessing. Apart from this, the servers also store the images in reliably higher ratio of memory and space. On demand from national or regional servers, the data is optimized and forwarded. These set of data is considered as internet data.

In our proposed work we have considered a demanded earth aerial pictorial data. These main servers are connected with inland servers and also proceed with connection with other primary servers. On request, the data is flowed from main server to inland servers via internet and thus we achieve the input data sets. Future, the samples are connected and stored into the regional servers and databases. The acquired images for Hadoop node is fetched from regional servers in our proposed system. Apart from storage, the data pre-processing and filtering is performed in this stage. Data cleaning is required, as acquired data is accompanied with other relevant and irrelevant attribute set.

Finally we move towards, MapReduce based data reduction and monitoring under a scalable and most efficient time of processing. The data acquired here is considered as a primary asset for processing and achieving image redundancy. The system also aims to focus on time consummation during processing huge data sets in Hadoop environment and Core Java environment. The proposed big data architecture of the remote data sensing has been divided into 3 parts 1) RSDU (Remote Sensing Big Data Acquisition Unit), 2) DPU (Data Processing Unit) and 3) DADU (Data Analysis and Decision Unit). The working of these 3 units is described below.

A. RSDU (Remote Sensing Big Data Acquisition Unit)

Remote sensing encourages the growth of observatory system of the earth as cost efficient parallel data acquisition system to fulfill certain computational requirement. For effectively analyzing big data there is a need of the parallel processing to process the big data in a proficient way. Therefore, the proposed technique i.e. RSDU (Remote Sensing Big Data Acquisition Unit) is presented in the architecture of remote sensing big data that gathers the data from various satellite from the globe. There is a possibility that raw data received can be distorted by various atmospheric gases and the dust particle.

We assume that the distorted or the erroneous data, satellite can correct. However, the remote sensing satellite uses the

algorithm Doppler or SPECAN to make the raw data into the image format. The data is sent for further processing to the earth base station by direct communication link. In these two types of data processing takes place one is offline data processing and another one is the real time data processing. In offline data processing, data will be transferred for storage, to the data center by the earth base station. This data is used for the future analysis. In the real time data, the data is directly sent to the FLBS (Filtration and Load Balancing Server.
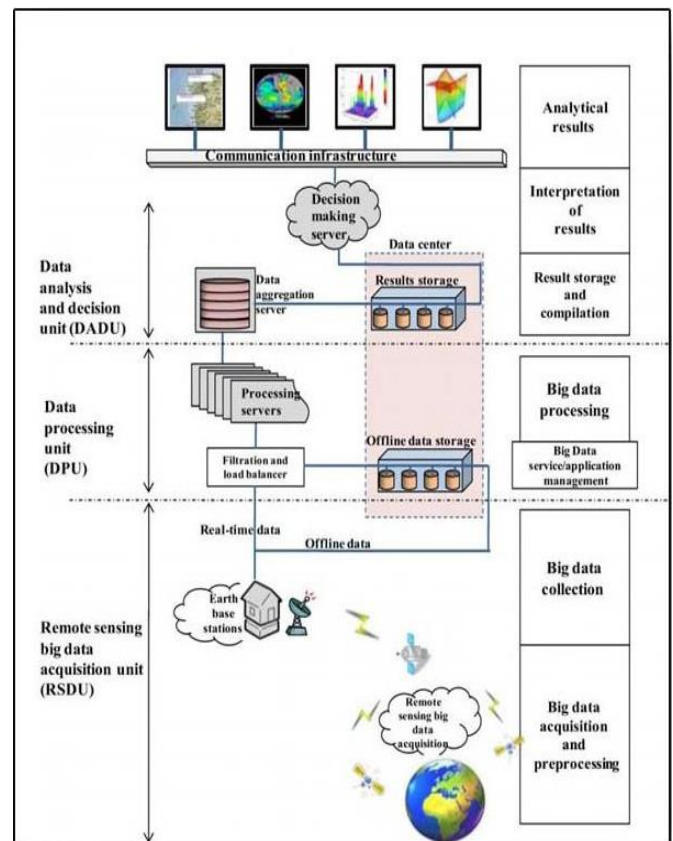

Fig 1: Remote Sensing Data Architecture

B. DPU (Data Processing Unit)

In DPU that is Data Processing Unit, it has two responsibilities. First, data need to be filtered by the filtration process. Second, balance the processing power by the load balancing server. Filtration recognizes or identifies the useful information, remaining data discarded of blocked. Hence, it improves the results of performance of the system. The load balancing server give the facility to divide the filtered data into parts and each part will be processed by the processing server. This load balancing and the filtration algorithm changes from analysis to analysis. Example, if there is a need for only temperature data and the sea wave, then the needed data is filtered out and it is divided into parts. Every processing server has its algorithm, to process the incoming segments of data from the filtration and the

load balancing server. The processing servers perform some measurements, statistical calculations and make other logical or mathematical operations to create the intermediate results from every segment of data. Since each processing servers executes the tasks in parallel and independently, hence the proposed system dramatically boosts the performance. The results obtained by every processing server are sent to further processing to the aggregation server for organization, compilation and storing.
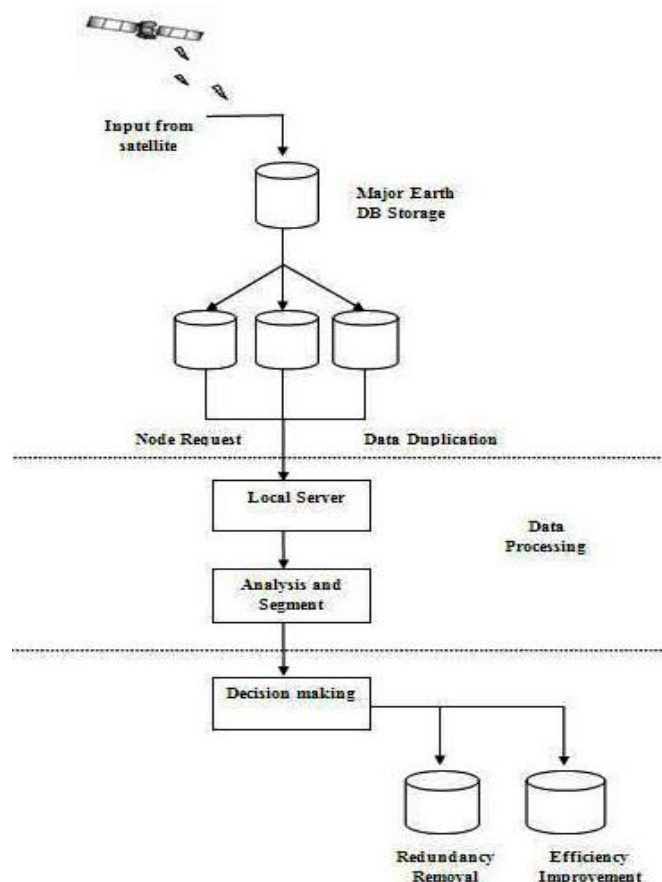

Fig 2:Analytical Architecture of Remote Sensing Big Data

C. Data Analysis and Decision Unit (DADU)

DADU contains three major portions, such as aggregation and compilation server, results storage server(s) and decision making server. When the results are able for compilation, the processing servers in DPU ship the partial outcome to the aggregation and compilation server, since the aggregated results are not in organized and compiled form. In the proposed architecture, aggregation and compilation server is supported by various algorithms that compile, organize, store, and transmit the results. Again, the algorithm varies from requirement to requirement and depends on the analysis needs.

Aggregation server stores the compiled and organized results into the result's storage with the intention that any server can use it as it can process at any time. The aggregation server also sends the same copy of that result to the decision-making server to process that result for making decision. The decision-making server is supported by the decision algorithms, which inquire different things from the result, and then make various decisions (e.g., in our analysis, we analyze land, sea, and ice, whereas other finding such as fire, storms, Tsunami, earthquake can also be found). The decision algorithm must be strong and correct enough that efficiently produce results to discover hidden things and make decisions. The decision part of the architecture is significant since any small error in decision-making can degrade the efficiency of the whole analysis.
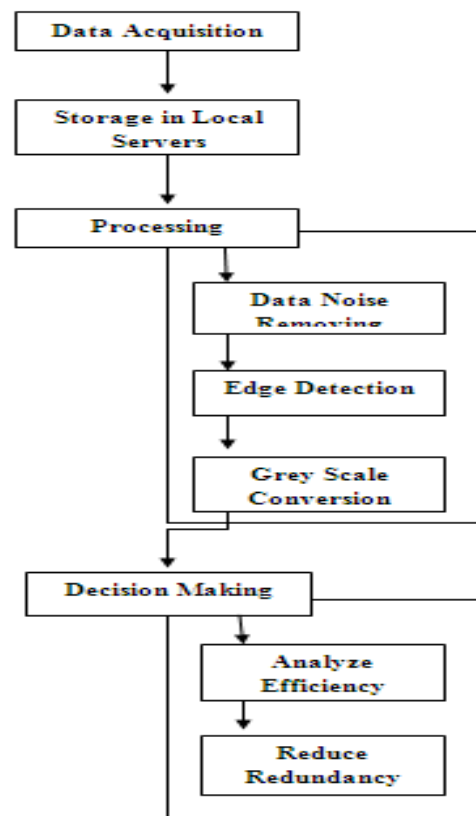

Fig 3: Flowchart - Remote Sensing Big Data Architecture

DADU finally displays or broadcasts the decisions, so that any application can utilize those decisions at real time to make their development. The applications can be any business software, general purpose community software, or other social networks that need those findings (i.e decision making). The self-explanatory flowchart supporting the working of the proposed architecture is depicted in Fig 3.

The system comprises of a planned server getting to of data images from satellite and is fetched on request. The data acquired is stored in local servers and subsequently the

request is processed in these servers for getting the data. The remote servers contain data greater than regular size thus processing time and storage is considerably high and thus we propose the system to process such complicated images in a simpler manner under Hadoop Cluster. Primarily the data is acquired and cleaned under Hadoop environment. The pre-processing step entails information pre-processing with information aligning and refining. The alignment of data is also carried out with filling an unfilled attributes. This is considered to be the primary processing step. In future steps, the data is analyzed and processed under filtering algorithm, the MapReduce operation is also performed in this step. The data collected is now stored offline under local regional servers.

5. RESULTS

Decision making and analysis is considered to be most difficult and standout step performed in our proposed system and thus fetches generic data sets for acquiring data analysis and controlling. The system also helps the users to achieve image redundancy optimization from regular and trivial storage systems. The system also contributes in this step to identify a distinct difference in processing a time gap from regular JAVA and Hadoop clustering. The major contribution of this system application is to achieve reliable results on selecting Hadoop clustering environment v/s java environment in terms of data processing and efficiency matching. The application also makes use of high data processing record and thus converts the given input data sets into gray images and extracts edges for achieving a secure data redundancy.

6. CONCLUSION

In this paper, we discussed the proposed a big data architecture for remote sensing applications. The proposed architecture is designed in such a way that, it can analyze both the offline as well as the real time data in an efficient way. In this paper, we discuss architecture for real-time Big Data analysis for remote sensing application. The Remote sensing Big Data architecture efficiently processed and analyzed real-time and offline remote sensing Big Data for decision-making. The proposed architecture contains three major units, such as 1)RSDU 2)DPU and 3)DADU. These units implement algorithms for each level of the architecture depending on the required analysis. The architecture of real-time big data is non specific (application independent) that is utilized for a remote sensing Big Data analysis. Besides, the capacities of filtering, dividing, and parallel processing of only useful information are performed by disposing of all other additional information. These processes make a better choice for real-time remote sensing Big Data analysis. The Remote Sensing Big Data architecture invites researchers and organizations for any type of remote sensory Big Data analysis by developing algorithms for each level of the architecture relying upon their analysis requirement.

7. FUTURE ENHANCEMENT

In future, we are planning to extend the proposed architecture to make it suitable for Big Data analysis for all applications, e.g., sensors and social networking. We are also planning to use the proposed architecture to perform complex analysis on earth observatory data for decision making at real-time, such as earthquake prediction, Tsunami prediction and fire detection.

8. REFERENCES

[1]Hung and Wn Ji, "Real Time Big Data Analytical Architecture for Remote Sensing Application" 2015.

[2] J. Cohan, B. Dolen, M. Dunlep, J. M. Hellerstein, and C. Welton, "Mad skills: New analysis practices for Big Data", PVLDB, vol. 2, no. 2, pp. 1481–1492, 2009.

[3] D. Agrawal, S. Das, and A. E. Abbadi, "BigData and cloud-computing: Current state and future opportunities", in ProcIntConf. Extending Database Technol. (EDBT), 2011, pp. 530–533.

[4] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters".

[5] K. Michael and K. W. Miller, "Big Data: New opportunities and new challenges [guest editors' introduction],pp. 22–24, Jun. 2013.

[6]C. Eaton, D. Deraos, T. Deuutsch, 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

[7] R. A. Dugane and A. B. Raut, "A survey on Big Data in real-time," Int. J. Recent Innov. Trends Comput. Commun., vol. 2, no. 4, pp. 794– 797, Apr. 2014.

[8] E. Chrisphe, J. Michel, and J. Inglada, "Remote sensing processing: From multicore to GPU".

[9] Yan Ma, Haiping wu, Lijhe wang, Bormon Huang, Rajiv Ranjan, Albert Zonmaya, wei jie, "Remote sensing big data computing: Challenges and Opportunities," Article in press., October 2014.

[10] A. Paul, K. Bharanitharan, and J.-F. Wang, "Region similarity based edge detection for motion estimation in H.264/AVC", Jan. 2010.

[11] A.-C. Tsai, A. Paul, J.-C. Wang, and J.-F. Wang, "Intensity gradient technique for efficient intra prediction in H.264/AVC", May 2008.

[12] R. A. Dugane and A. B. Raut, "A survey on Big Data in real-time," Int. J. Recent Innov. Trends Comput. Commun., vol. 2, no. 4, pp. 794– 797, Apr. 2014.