International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 23 Issue 7 –OCTOBER 2016.

High Dimensional Big Data Analytical Architecture for Real-Time Remote Sensing Application

T.Swathi[#1] and A.Poorna Chandra Reddy[*2]

[#] *Christu Jyoti Institute of Technology and Science, Jangaon,Warangal(Dist),Telagana(State) India*

[*] *Assistant Professor, Computer Science Engineering Department, Christu Jyothi Institute of Technology and Science, Jangaon, Warangal(Dist), Telagana(State) India*

***Abstract*— Big Data is the new experience curve in the new economy driven by data with high volume, velocity and variety. The real-time remote sensing Big Data seems at first, and extracting the useful information in an effective manner leads a system toward a massive computational challenges, such as to analyze, aggregate, and store, where data are remotely gathered. Keeping in view the over mentioned factors; there is a need for designing a system architecture that welcomes both real-time, as well as offline data processing. Therefore, in this paper, we discuss real-time Big Data analytical architecture for remote sensing satellite application. In Architecture contains three main units, such as 1) remote sensing Big Data acquisition unit (RSDU); 2) data processing unit (DPU); 3) data analysis decision unit (DADU). Firstly, RSDU gathers data from the satellite and transmit this data to the Base Station, where initial processing takes place. Later DPU plays an essential role in architecture for efficient processing of real-time Big Data by providing filtration, load balancing, and parallel processing. Then DADU is responsible for compilation, storage of the results, and generation of decision based on the results received from DPU. The architecture has the capability of dividing, load balancing, and parallel processing of only useful data. Thus, it results in efficiently analyzing real-time remote sensing Big Data using earth observatory system. Finally, a detailed analysis of remotely sensed earth observatory Big Data for land and sea area are provided using Hadoop.**

***Index Terms*—Big data, RSDU, DADU, Hadoop**

I. INTRODUCTION

Technology has played the roles of enabler and driver in the evolution of the economies spanning the eras characterized by agriculture, manufacturing, service, and knowledge assets. Alongside the change of the economies, technology has evolved across the mainframe computer, the PC, client-server computing, the Internet, cloud computing, mobile computing and social networking. Big Data emerges as the latest stage of the evolution that combines three trends in technology, which Minelli, Chambers, and Dhiraj (2013) described as the three perfect storms: computing, data and convergence. The computing storm results from the exponential growth of processing power as predicted by Moore's Law, mobile computing, social network, and cloud computing.

The data storm results from the accessibility of data with high volume, velocity and variety. The convergence storm results from the availability of open-source technology and commodity hardware. Big Data from the technology standpoint are datasets that require beyond the currently available technological capacity. From the business standpoint, they represent a new strategy of creating actionable business insights enabling organizations to sense and respond in a rapidly changing environment. The impact of Big Data is felt across many sectors and industries.

King and Rosenbush (2013) reported that Sears Holding Corporation and Wal-Mart Stores, Inc., use Big Data database for marketing efforts, and Chevron Corporation uses them to process seismic data in the search for new reserves of oil and gas. Winslow (2013) reported the use of Big Data to collect data on the care of hundreds of thousands of cancer patients and use it to help guide treatment of other patients across the healthcare system. Mahrt and Scharkow (2013) described the value of Big Data in digital media research, where the data rush through social media promises new insights about consumers' behavior.

The impact of real-time social media was felt when the Dow was down over a hundred and forty points in early trading on April 23, 2013 after the Associated Press reported the false tweet about an attack at the White House (Associated Press, 2013). Government has taken notice of Big Data as well and has announced a Big Data Research and Development initiative to improve the ability to extract knowledge and insights from large and complex collections of digital data to help solve some of the Nation's most pressing challenges spanning across concerns in science and engineering, healthcare and national security (Executive Office of the President, 2012). Harbert (2013) postulated that Big Data helps to create big career opportunities that include data scientists, data architects, data visualizers, data change agents, data engineers and operators

To address the aforesaid needs, a remote sensing Big Data analytical architecture [1], this is used to analyze real time, as well as offline data. At first, the data are remotely preprocessed, which is then readable by the machines. Afterward, this meaningful information is delivered to the Earth Base Station for further data processing. Earth Base Station (EBS) performs two types of processing, such as processing of real-time and offline data. In case of the offline data, the data are transmitted to offline data-storage device. The incorporation of offline data-storage device helps in later usage of the data, whereas the real-time data is directly transfer to the filtration and load balancer server, where filtration algorithm is employed, which drag out the meaningful information from the Big Data. On the other hand, the load balancer balances the processing power by equal distribution of the real-time data to the servers. The filtration and load-balancing server not only filters and balances the load, but it is also used to enhance the system efficiency.

II. BIG DATA ANALYTICS

Chen, Chiang, and Storey (2012) provided a classification of business intelligence and analytics (BI&A) into three categories. BI&A 1.0 is characterized by DBMS-based and structured content. It utilizes traditional analytic tools via data warehousing, ETL, OLAP and data mining. BI&A 2.0 is characterized by Web-based and unstructured content. It utilizes tools in information retrieval, opinion mining, question answering, Web analytics, social media analytics, social network analysis, and spatial-temporal analysis. BI&A 3.0 is characterized by mobile and sensor-based content. It utilizes tools in location-awareness analysis, person-centered analysis, context-relevant analysis, and mobile visualization and HCI. BI&A 2.0 and 3.0 would require a platform that can handle the huge volume, velocity and variety of data. The Big Data analytics architecture described below utilizes the massively parallel, distributed storage and processing framework as provided by Hadoop HDFS and MapReduce. As opposed to some belief that Big Data has pronounced the obsolescence of data warehousing, it remains a viable technology for Big Data analytics of huge volume of structured data. Furthermore, there is synergy between data warehousing and the Hapdoop type Big Data architecture. Unstructured data from sensors, M2M devices, social media and Web applications can be stored in Hadoop and be MapReduced later for meaningful insight (Sathi, 2012). MapReduced data can then be integrated with the data warehouse for further analytic processing. Conversely, data warehouse can be a data source for complex Hadoop jobs, simultaneously leveraging the massively parallel capabilities of two systems (Awadallah& Graham, 2011). Real-time location data from GPS or smartphones can be combined with historic data from the data warehouse to provide realtime insight for marketers to promote products targeted to the individual customer based on real-time location data and customer profile. Figure 3 illustrates an architecture for Big Data analytics.

III. RELATED WORK

Work related to our project includes research in cloud computing, Big Data, Map reduce. Author of [1] explains about Big Data and cloud computing terms. Scalable database management systems (DBMS)-both for update intensive application workloads as well as decision support systems for descriptive and deep analytics are a critical part of the cloud infrastructure and play an important role in ensuring the smooth transition of applications from the traditional enterprise infrastructures to next generation cloud infrastructures. Author of [3] explain about map reduce. MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. Author of [4] explains about starfish. Timely and cost-effective analytics over “Big Data ” is now a key ingredient for success in many businesses, scientific and engineering disciplines, and government endeavors. The Hadoop software stack which consists of an extensible MapReduce execution engine, pluggable distributed storage engines, and a range of procedural to declarative interfaces—is a popular choice for big data analytics. In this paper we fetch the twits on the server then we preprocess those twits using proposed architecture.

IV. PROPOSED MODEL

We discuss remote sensing Big Data architecture [1] to analyze the Big Data in an effective manner as shown in Fig. 1. Fig. 1 describes n number of satellites that acquire the earth observatory Big Data images with sensors or conventional cameras through which sceneries are recorded using radiations. Appropriate techniques are applied to process and interpret remote sensing imagery for the aim of producing conventional maps, thematic maps, resource surveys, etc. It assumes that the data are huge in nature and difficult to handle for a single server. The data are continuously arriving from a satellite with high speed. Hence, special algorithms are needed to process, analyze, and make a decision from that Big Data. Here, we analyze remote sensing data for finding land, sea, or ice area. In remote sensing Big Data architecture contain mainly three units, i.e., 1) remote sensing data acquisition unit (RSDU); 2) data processing unit (DPU); and 3) data analysis and decision unit (DADU). The functionalities and working of this unit are described as below. Fig.1 Remote Sensing Data Architecture A. Remote Sensing Big Data Acquisition Unit (RSDU) The RSDU is introduced in the remote sensing Big Data architecture that gathers the data from various satellites around the globe as shown in Fig. 2 [14]. It is possible that the received raw data are distorted by scattering and absorption by various atmospheric gasses and dust particles. We assume that the satellite can correct the erroneous data. However, to bring the raw data into image format, the remote sensing satellite uses Doppler or SPECAN algorithms [13]. For efficient data analysis, remote sensing satellite preprocesses data under many situations to integrate the data from different sources, which not only decreases storage cost, but also improves analysis accuracy. Some relational data preprocessing techniques are data integration, data cleaning, and redundancy elimination [15]. After preprocessing phase, the assemble data are delivered to a ground station using downlink channel. This transmission is directly or via relay satellite with a special tracking antenna and communication

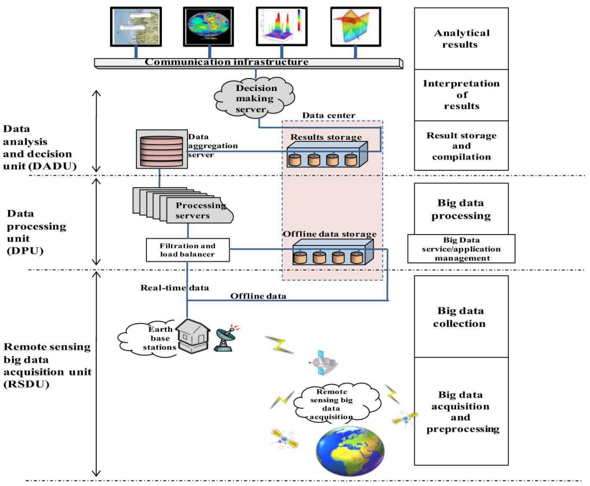

Fig 1: Remote sensing Big Data architecture.

A. *Data Processing Unit In data processing unit (DPU),*

The filtration and load balancer server have two basic responsibilities, such as filtration of data and load balancing of processing power. Filtration identifies the useful data for analysis since it only allows meaningful information, whereas the rest of the data are blocked and are discarded. Hence, it results in enhancing the performance of the whole proposed system. Apparently, the load-balancing part of the server provides the facility of dividing the whole filtered data into parts and assign them to various processing servers. The filtration and load-balancing algorithm varies from analysis to analysis. Each processing server has its algorithm implementation for processing incoming segment of data from FLBS. Each processing server makes statistical calculations, any measurements, and performs other mathematical or logical tasks to generate intermediate results against each segment of data. Since these servers perform function individually and in parallel, the performance of system is dramatically enhanced, and the results against each segment are generated in real time. The results generated by each server are then sent to the aggregation server for compilation, organization, and storing for further processing.

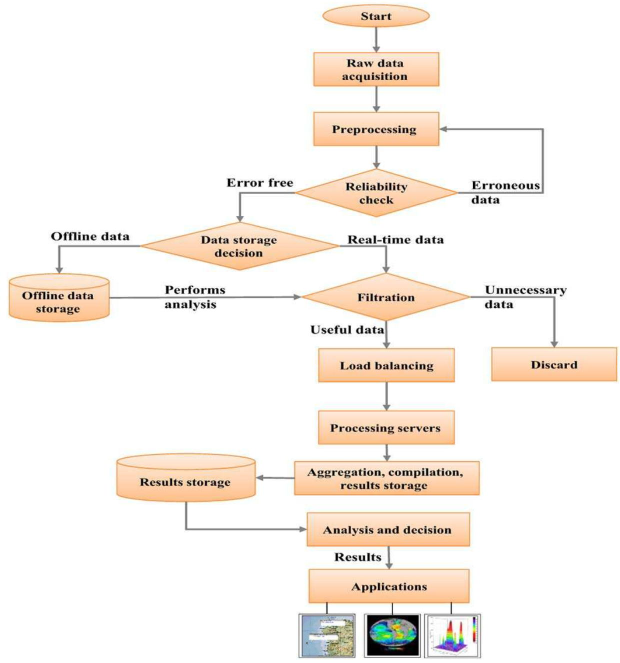

Fig 2: Flow Chart

B. *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 ready for compilation, the processing servers in DPU send the partial results to the aggregation and compilation server, since the aggregated results are not in organized and compiled form. Therefore, there is a need to aggregate the related results and organized them into a proper form for further processing and to store them. 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. Fig.2 Flowchart of Remote Sensing Big Data Architecture 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 algorithm, which inquires different things from the result, and then makes 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. 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).

V. METHODOLOGY

Problems in Project development, especially in the software related field are due to its complex nature, could often encounter many unanticipated problems, resulting in the projects falling behind on deadlines, exceeding budgets and result in sub-standard products. Although these problems cannot be completely eliminated, they can however be controlled by applying Risk Management methods. This can help us to deal with problems before they occur. Organizations who implement risk management procedures and techniques will have greater control over the overall management of the project. By analyzing five of the most commonly used methods of risk management, conclusions will be drawn regarding the effectiveness of each method

VI. ANALYSIS AND DISCUSSION

Understanding environment requires massive amount of data collected from various sources, such as remote access satellite observing earth characteristics [measurement data set (MDS) of satellite data such as images], sensors monitoring air and water quality, metrological circumstances, and proportion of CO2 and other gases in air, and so on. Using the architecture for offline as well online traffic, we perform a simple analysis on remote sensing earth observatory data. We assume that the data are big in nature and difficult to handle for a single server. The data are continuously coming from a satellite with high speed. Hence, special algorithms are needed to process, analyse, and make a decision from that Big Data. Here, in this section, we analyse remote sensing data for finding land, sea, or ice area. We have used the proposed architecture to perform analysis and proposed an algorithm for making decision Hadoop gives the facility of parallel, high-performance computing using a large number of servers. Therefore, it is suitable for analysing a large amount of remote sensory image data. The architecture uses a similar mechanism for load balancing; hence, preference is given to Hadoop for sophisticated analysis, algorithm development, and testing.

VII. CONCLUSION

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 is 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 is generic (application independent) that is used for any type of remote sensing Big Data analysis. Furthermore, the capabilities of filtering, dividing, and parallel processing of only useful information are performed by discarding all other extra data. These processes make a better choice for real-time remote sensing Big Data analysis. The Remote Sensing Big Data architecture welcomes researchers and organizations for any type of remote sensory Big Data analysis by developing algorithms for each level of the architecture depending on their analysis requirement. For future work, 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, fire detection, etc.

.

REFERENCES

[1] Muhammad MazharUllahRathore, Anand Paul, Bo-Wei Chen, Bormin Huang, and Wen Ji, "Real-Time Big Data Analytical Architecture for Remote Sensing Application," IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING ,2015.

[2] J. Cohen, B. Dolan, M. Dunlap, 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, "Big Data and cloud computing: Current state and future opportunities," in Proc. Int. Conf. Extending Database Technol. (EDBT), 2011, pp. 530–533.

[4] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/texarchive/macros/latex/contrib/supported/IEEEtran/

[5] 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.

[6] "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.

[7] R. D. Schneider, Hadoop for Dummies Special Edition. Hoboken, NJ, USA: Wiley,2012

[8] J. Shi, J. Wu, A. Paul, L. Jiao, and M. Gong, "Change detection in synthetic aperture radar image based on fuzzy active contour models and genetic algorithms," Math. Prob. Eng., vol. 2014, 15 pp., Apr. 2014.

[9] A. Paul, J. Wu, J.-F. Yang, and J. Jeong, "Gradient-based edge detection for motion estimation in H.264/AVC," IET Image Process., vol. 5, no. 4, pp. 323–327, Jun. 2011.

[10] A. Paul, K. Bharanitharan, and J.-F. Wang, "Region similarity based edge detection for motion estimation in H.264/AVC," IEICE Electron. Express, vol. 7, no. 2, pp. 47–52, 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," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 5, pp. 694–698, May 2008.

[12] Muhammad MazharUllahRathore, Anand Paul, "An efficient divide-and-conquer approach for big data analytics in machine-to-machine communication" The School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, Republic of Korea Comput. Intell. Comput. Res. (ICCIC), 2013.

[13] EnviSat, A. S. A. R. Product Handbook, European Space Agency, Issue 2.2, Feb., 2007.

[14] European Space Agency. (Oct. 14, 2014). [2312] [Online]. Available: https://earth.esa.int/

[15] L. Angrisani, M. D'Arco, G. Ianniello, and M. Vadursi, "An efficient pre-processing scheme to enhance resolution in band-pass signals acquisition," IEEE Transa. Intrum. Meas., vol 61, no. 11, pp. 2932–2940, Nov. 2012.