International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 2, Issue 3 (March '2016)

ENHANCING THE PERFORMANCE OF FILTERING USING CANNY EDGE DETECTION ALGORITHM FOR REAL TIME BIG DATA WITH HADOOP

K. Akshaya,
Department of Computer science and Engineering,
Dhanalakshmi Srinivasan Institute of Technology,
Trichy, India.

K.Gangatharan,
Department of Computer science and Engineering,
Dhanalakshmi Srinivasan Institute of Technology,
Trichy, India.

R.Vishnu Priya,
Department of Computer science and Engineering,
Dhanalakshmi Srinivasan Institute of Technology,
Trichy, India.

N. Suresh Kumar,
Assistant Profressor,
Department of Computer science and Engineering,
Dhanalakshmi Srinivasan Institute of Technology,
Trichy, India.

Abstract: TheProposed architectonics comprises threemain units, such as Remote Sensing Big Data Acquisition Unit (RSDU), Data Processing Unit (DPU) and Data Analysis Decision Unit (DADU). First, RSDU acquires datafrom the satellite and dispatchesacquired data to the Base Station, where initial process takes place. Second, DPU plays a vital role in architecture, for efficient processing of real-time Big Data by providing filtration, load balancing, and parallel processing. Finally, DADU is the uppermosttier of the proposed scheme, which is responsible for compilation, storage of the outcome, and reproduction of decision upon the outcome received from DPU. Thus, it results in analyzing the actual time, for remote sensing Big Data using globe observatory system. Furthermore, the proficiency of storing incoming data to implement offline analysis on largely stored dumps, when required. Eventually, a detailed analysis of remotely sensed globe observatory Big Data for land and sea area are provided using hadoop and Canny edge detection algorithmwhich uses a multi-stage algorithm to detect wide range of edges in images.

Keywords: Remote Sensing Big Data Acquisition Unit (RSDU), Data Processing Unit (DPU), Data Analysis Decision Unit (DADU),Big data, Hadoop, Canny edge detection algorithm.

I. INTRODUCTION

Big Data classifies specialized types of data sets comprising formless data, which dwell in data tier of technical computing applications and the Web. The data stored in the underlying tier of all these technical computing application scenarios have some literal uniqueness in common, such asFirst, largescale data which refers size and data warehouse; Second, Scalability issues which refers application's likely to be running on large scale (e.g., Big Data); Third, Extraction Transformation Loading (ETL) method from low, raw data to well thought-out data up to certain extent; and Fourth, Development of uncomplicated interpretable analytical over Big Data warehouses with a view to deliver an intelligent and momentous knowledge for them .

II. EXISTING SYSTEM

Big Data analysis is somehow a challenging task than locating, identifying, understanding, and citing data. Having a large-scale data, all of this has to happen in a mechanized manner since it requires diverse data structure as well as semantics to be eloquent in computer- legible format. However, by analyzing simple data having one data set, a system is required of how to design a database. There might be alternative ways to store all of the same information. In which, the noticed design might have angain over others for certain process and possible drawbacks for some other purposes. In order tospot these needs, diverse analytical platforms have been provided by relational databases vendors. These platforms come in diverse shapes from software only toanalytical services that run in third-party hosted atmosphere. In remote connection networks, where the data source such as sensors can produce an overwhelming amount of data. In the first step, i.e., data acquisition, in which much of the data are of no interest that can be compressed or filtered by orders of magnitude. By viewing to use such filters, they do not discard useful information. For instance, in attention of new statements, is it adequate to keep that information that is noticed with the company name? Alternatively, is it necessary that the entire report?or simply a small piece around the noticed name? The second challenge is by default generation of accurate metadata that describe the composition of data and the way it was analyzedand collected. This kind of metadata is hard to analyze since to know the source for every individual data in remote connection. Normally, the data collected from remote areas are not in a format, ready for analysis. Therefore, the second step point outs us to data extraction, which drags out the useful information from the underlying sources and delivers in a structured format which ready for analysis. For instance, the data set is reduced to single-class tag to facilitate analysis, although the first stepthat is used to think about Big Data as always describing the actual process. However, this is far away from reality; sometimes it deals with inaccurate data too, or some of the data might be unspecified.

III. PROPOSED SYSTEM

Remote sensing Big Data architecture to analyze the Big Data in an efficient manner . Delineates n number of satellites that pick up the globe observatory Big Data images with sensors or conventional cameras through which landscape are recorded using radiations. Whereas specialmethods are applied to process and interpret remote sensing imagery for the purpose of producing resource surveys,conventional maps, thematic maps, etc. In our architecture, Remote sensing Big Data architecture into three parts. The data are remotely preprocessed, which is then readable by the machines. Afterward, this relevant information is delivered to the Globe Base Station for further data processing. Globe Base Station (GBS) performs two types of processing, such as real-time(online) and offline data. In case of the offline data, the data are delivered to offline data-storage device. The incorporation of offline data-storage device helps in coming afterconsumption of the data, In which the real-time data is transfer to the filtration and load balancer server, where filtration algorithm is working,which drag out the relevantinformation from the Big Data. On the other hand, the load balancer stabilizes the processing power by equal distribution of the real-time data to the servers. The load-balancing andfiltration server not only stabilizeand filters the load, but it is also used to enhance the system efficiency. Hadoopgrasp a gap in the market by effectively providingandstoring computational capabilities over substantial number of data. Whereas distributed system offers a way to parallelize and execute programs on a cluster.

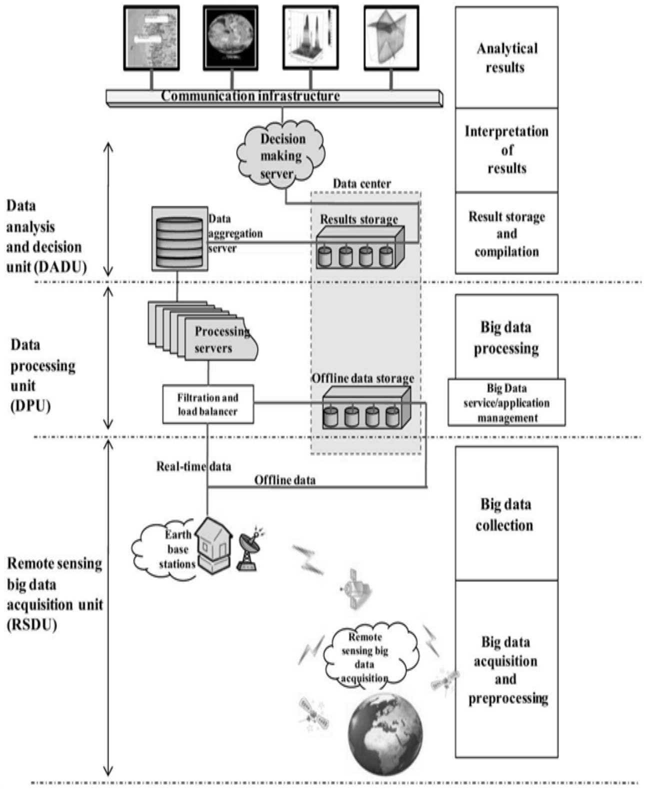

Fig. 1.Proposed Architecture

You've most likely come across Hadoop as it's been adopted by technology giants like Twitter, Yahoo,and Facebookto address their big data needs, and it's making inroads across all industrial sectors.

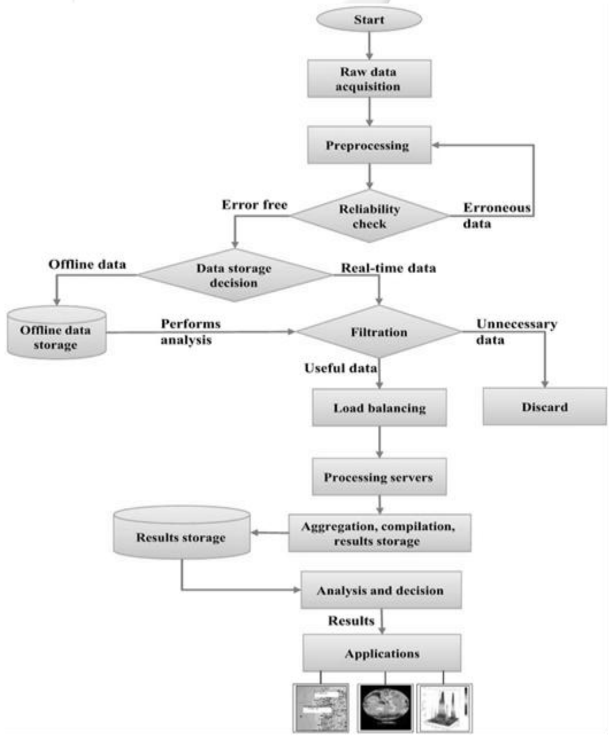

Fig. 2. Flowchart of the remote sensing Big Data architecture

A. Remote Sensing Big Data Acquisition Unit (RSDU)

It promotes the expansion of globe observatory system as cost-effective parallel data acquisition systemto satisfy precise computational requirements. TheSpace Science and GlobeSociety accepted this solution as the a level of quality for parallel processing in this precise context As satellite instruments for Globe observation integrated much sophisticated qualities for improving Big Data acquisition, soon it was recognized that conventional data processing technologies could not hand oversufficient power for processing such kind of data.

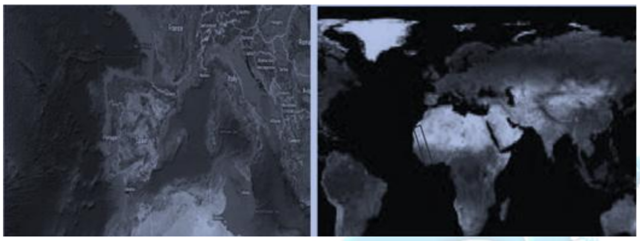

Fig. 3.Satellite view of Land and Sea Area

Therefore, the need for parallel processing of the huge volume of data was needed, which could efficiently analyze the Big Data. For that reason, the RSDU is introduced in the remote sensing Big Data architecture that gathers the data from diverse satellites around the globe it is possible that the received raw data are distorted by scattering and absorption by diverse dust particles andatmospheric gasses. Consider the satellite can correct the erroneous data. However, to make the data into image format, the remote sensing satellite uses Doppler or SPECAN For effective data analysis,remote sensing satellite preprocess data beneath many situations to integrate the data from different sources, which is not only reduces storage cost, but also improves analysis accuracy. Some relational data preprocessing techniques are data cleaning, data integration, and redundancy eliminationafter preprocessing phase, the collected data are transmitted to a landscape using downlink channel.This transportation is directly or via relay satellite with an appropriate tracking of communication link and antenna in a wireless atmosphere. The data must be corrected in different methods to remove distortions.Because the motion of the platform related to the globe, globe curvature,platform attitude,variations in sensor characteristics,no uniformity of illumination, etc.,For further process direct communication link is used when the data is transmitted to Globe Base Station. Consider the data processing procedure into two steps, such as offline Big Data processing andreal-time Big Data processing. In the case of offline data processing, The Globe Base Station transferand save the data in data center. This data is then used for future analyses. However, The data are directly transmitted to the filtration and load balancingserver in real-time process, after all storing of incoming real-time data diminishes the performance of real-time processing.

B. Data Processing Unit (DPU)

In data processing unit, the load balancer andfiltration server have two basic responsibilities, such as load balancing of processing power and filtration of data. Filtration identifies the beneficial data for analysis so it only permitsbeneficial information, whereas the inactive of the data are blocked and are discarded. Hence, itelevatesthe performance of whole proposed system. Apparently, the load-balancing is one part of the server provides the provision of split up the whole part and filtered data into parts thenassigns them to various processing servers. The filtration and load-balancing algorithm differ from analysis to analysis; e.g., if there is only a need for analysis of temperature and sea wave of data, the measurement of these described data is filtered out, and is segmented into parts. Each server has its algorithm implementationfor processing incoming segment of data from FLBS. Eachserver makes measurements,performs other mathematical or logical tasks and statistical calculations to generate intermediate results opposite to each segment of data. Since those servers are performing tasksparallel andindependently, the performance of proposed system is dramatically enhanced, and the results opposite to each segment are generated in real time. The results generated by each server are then sent to the aggregation server for organization, compilation, and storing for further processing.

Canny edge detection algorithmis an better edge detection performance because it satisfies three main criteria which are low error rate, good localization and minimal response. In this paper, a mechanism to implement the Cannyedge detection algorithm at block level with enhanced edge detection performance is proposed. By directly applying the original frame-level Canny algorithm at block level leads to more number of edges in smooth regions and to loss of important edges in highlydetailed regions since the original Canny algorithm computes the high and low thresholds based on the frame-level statistics.

C. *Data Analysis Decision Unit (DADU)*

DADU contains three major portions, such as compilation and aggregationserver, decisionmaking server and results storage server(s). When the results are ready forcompilation, the processing servers in DPU transmit the limited results to the compilation andaggregation server, since the aggregated results are not in compiled andorganized form. Hence, there is a necessity to aggregate the relevant results and organized them into a proper form.In further it is stored. In the proposed architecture, compilationand aggregation server are assist by various algorithms that organize, compile,transmit and storethe results. Again, the algorithm diverse from every requirement and it depends on the analysis. Aggregation server stores the organized and compiled results into the result's storage. So server can use it as it can process at any time. The aggregation server also transmits the same copy of that result to the decision-making server to process that result for making decision. Whereas decision making algorithmassistthe decision-making serverwhich inquire different things from the result and then make various decisions. The decision algorithm must be correctand strong enough that efficiently produce results to discover concealed things and make decisions. The decision part of the architecture is significant since any small error in decision-making can diminish the capability of the whole analysis. EventuallyDADU displays or broadcasts the decisions, so that any application can deploy those decisions at real time to make their progress. The applications can be general purpose community software, or other social networks, any business software.Thatnecessityof those findings are helps self-explanatory flowchart supporting the working of the proposed architecture.

IV. CONCLUSION

The proposed architecture efficiently processed and analyzed real-time and offlineremote sensing Big Data for decision-making.It is composed of three major units, such as 1) RSDU; 2) DPU; and 3) DADU. These units implement algorithms foreach 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 algorithms proposed in this paper for each unit and subunits are used to analyze remote sensing data sets, which helps in better understanding of land and sea area. The proposed 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, the proposed architecture to make it compatible for Big Data analysis for all applications, e.g., sensors and social networking.Further the proposed architecture to perform complex analysis on earth observatory data for decision making at realtime, such as earthquake prediction, Tsunami prediction, fire detection, etc

V. REFERENCES

[1]. A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers," Concurrency and Computation: Practice and Experience(CCPE), 2012, DOI: 10.1002/cpe.1867, (in press).

[2]. N. Bobroff, A. Kochut, and K. Beaty, "Dynamic placement of virtual machines for managing SLA violations," in Proc. of the 10th IFIP/IEEE Intl. Symp. on Integrated Network Management (IM), 2007, pp. 119–128.

[3]. Christian Strack "Performance and Power Management for Cloud Infrastructures" Philipps-University Marburg,

Germany -Department of Mathematics and Computer Science, Distributed Systems Group.

[4]. Gartner, Inc.,Gartner estimates ICT industry accounts for 2 percentof global CO2 emissions. Gartner Press Release (April 2007).

[5]. D. Gmach, J. Rolia, L. Cherkasova, G. Belrose, T. Turicchi, andA. Kemper, "An integrated approach to resource pool management:Policies, efficiency and quality metrics," in Proc. of the 38thIEEE Intl. Conf. on Dependable Systems and Networks (DSN), 2008,pp. 326–335.

[6]. D. Gmach, J. Rolia, L. Cherkasova, and A. Kemper, "Resourcepool management: Reactive versus proactive or lets be friends,"Computer Networks, vol. 53, no. 17, pp. 2905–2922, 2009."A Iosup-2010" on the performance variability of production cloud services

[7]. J. Koomey,Growth in data center electricity use 2005 to 2010.Oakland, CA: Analytics Press, 2011.

[8]. S. Kumar, V. Talwar, V. Kumar, P. Ranganathan, and K. Schwan,"vManage: Loosely coupled platform and virtualization managementin data centers," in Proc. of the 6th Intl. Conf. on AutonomicComputing (ICAC), 2009, pp. 127–136.

[9]. " X Meng - 2010" Efficient Resource Provisioning in Compute Clouds via VM Multiplexing"

[10]. R. Nathuji and K. Schwan, "Virtualpower: Coordinated powermanagement in virtualized enterprise systems,"

[11]. "C Strack - 2012 "Performance and Power Management for Cloud Infrastructures

[12]. C Tang-2007" A Scalable Application Placement Controller for Enterprise Data Centers

[13]. A. Verma, P. Ahuja, and A. Neogi, "pMapper: Power and migration cost aware application placement in virtualized systems," inProc. of the 9th ACM/IFIP/USENIX Intl. Conf. on Middleware, 2008,pp. 243–264.

AUTHOR PROFILES

K. Akshaya
Dept. of Computer science andEngineering,
Dhanalakshmi Srinivasan Institute of Technology,Trichy, India.
Mail: atchu00221@gmail.com

K. Gangatharan
Dept. of Computer science andEngineering,
Dhanalakshmi Srinivasan Institute of Technology,Trichy, India.
Mail: gangacse004@gmail.com

R.VishnuPriya
Dept. of Computer science andEngineering,
Dhanalakshmi Srinivasan Institute of Technology,Trichy, India.
Mail: jananipriya1812@gmail.com

N. SURESH KUMAR M.E.,
Assistant Professor,
Dept. of Computer Science and Engineering,
Dhanalakshmi Srinivasan Institute of. Technology, Trichy, India.
Mob: +91 80567 44016
Mail: sureshkumar0707@gmail.com
sureshkumarcse2@gmail.com