Volume 3, Issue 6, June-2016, pp. 279-283 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

Effective Analysis of Data from Remote Sensing Application

[1]Santhosh Gowda T.R, [2]Vijay G.R

[1] Dept of CSE, SJCIT, Chickballapur
[2] Dept of CSE, SJCIT, Chickballapur
Santhoshgowda085@gmail.com, vijay.gcr@gmail.com

Abstract:- At present applications like Internet, mobile devices, social media, geospatial devices, sensors will generate massive volume of data.Processing and extracting the useful information in an efficient manner leads a system toward a major computational challenges, such as to analyze, aggregate, and store data. For these Big data analytical architecture is proposed. The architecture comprises 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). RSDU acquires data from the sensors and sends this data to the Base Station. DPU provides an efficient processing of Data by providing filtration, load balancing, and parallel processing. DADU is responsible for compilation, storage of the results, and generation of decision based on the results received from DPU.

Keywords— Big Data, remote sensing Big Data acquisition, data processing unit, data analysis decision unit.

1. INTRODUCTION

As moving data generators, human beings create data everyday[1]. We are all connected by sharing data from social networks, intelligent devices, etc. Remote sensing devices have been widely used to observe our planet from various perspectives and to make our lives easier. It is not exaggerated to say that the whole Earth has now been made digital. Therefore, the digitized Earth plus the moving data generators are the main actors for big data in remote sensing.

Remote sensing is generally defined as the technology of measuring the characteristics of an object or surface form a distance [2,3].The RS data are the earth observing data continuously obtaining from sensors, as well as some other data acquisition measurements. With the exponential growth of data amount and increasing degree of diversity and complexity, the remotely sensed data are regarded as RS ''Big Data''. However, since the whole idea of big data is still remaining relatively new, most of the start off efforts is focusing on the definition and discussion of the realm of the big data. Big data [4,5] occurs when a large collection of data sets whose volume and rate of data is at a scale that is far beyond the state-of-the-art systems and revolutionize the way of seeking solutions. This is also the case for the remote sensing and earth sciences domain to offer the definition of what RS ''Big Data'' really is. The RS ''Big Data'' not merely refers to the volume and velocity of data that outstrip the storage and computing capacity, but also the variety and complexity of the RS data.

Big Data analysis is somehow a challenging task for locating, identifying and understanding data [6]. 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 articulated in forms of computer-readable format. 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 such conditions, the mentioned design might have an advantage over others for certain process and possible drawbacks for some other purposes. In order to address these needs, various analytical platforms have been provided by relational databases vendors [7]. These platforms come in various shapes from software only to analytical services that run in third-party hosted environment.

In remote access networks, where the data source such as sensors can produce an overwhelming amount of raw data. We refer it to the first step, i.e., data acquisition, in which much of the data are of no interest that can be filtered or compressed by orders of magnitude. With a view to using such filters, they do not discard useful information. The second challenge is by default generation of accurate metadata that describe the composition of data and the way it was collected and analyzed. Such kind of metadata is hard to analyze since we may need to know the source for each data in remote access.

The data collected from remote areas are not in a format ready for analysis. Therefore, the second step refers us to data extraction, which drags out the useful information from the underlying sources and delivers it in a structured formation suitable for analysis. To address the aforementioned needs, this paper presents a Big Data analytical architecture, which is used to analyze real time, as well as offline data.

2. MOTIVATION FOR REMOTE SENSING BIG DATA ANALYTICS

The increase in the data rates generated on the digital universe is escalating exponentially. With a view in employing current tools and technologies to analyze and store, a massive volume of data are not up to the mark [8], since they are unable to extract required sample data sets. Therefore, we must design an architectural platform for analyzing both remote access real- time and offline data. When a business enterprise can pull-out all the useful information obtainable in the Big Data rather than a sample of its data set, in that case, it has an influential benefit over the market competitors. Big Data analytics helps us to gain insight and make better decisions. Therefore, with the intentions of using Big Data, modifications in paradigms are at utmost. To support our motivations, we have described some areas where Big Data can play an important role. 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. Through relating all the information drifting such as CO2 emanation, increase or decrease on greenhouse effects and temperature, can be found out.

In healthcare scenarios, medical practitioners gather massive volume of data about patients, medical history, medications, and other details. The above-mentioned data are accumulated in drug-manufacturing companies. The nature of these data is very complex, and sometimes the practitioners are unable to show a relationship with other information, which results in missing of important information. With a view in employing advance analytic techniques for organizing and extracting useful information from Big Data results in personalized medication, the advance Big Data analytic techniques give insight into hereditarily causes of the disease.

3. PHASES IN PROCESSING DATA

3.1 Data Acquisition and Recording

Big data does not arise out of a vacuum it recorded from some data generating source. For example consider our ability to sense and observe the world around us, from the heart rate of an elderly citizen, and presence of toxins in the air we breathe, to the planned square kilometers array telescope, which will produce up to 1 million terabytes of raw data per day. Similarly, scientific experiments and simulations can easily produce peta-bytes of data today.

The second big challenge is to automatically generate the right metadata to describe what data is recorded and how it is recorded and measured. For example, in scientific experiments, considerable detail as regarding specific experimental conditions and procedures maybe required to be able to interpret the results correctly, and it is important that such metadata be recorded with observational data.

3.2 Information Extraction and Cleaning

Frequently, the information collected will not be in a format ready for analysis. For example, consider the collection of electronic health records in a hospital, comprising transcribed dictations from several physicians, structured data from sensors and measurements(possibly with some associated uncertainty), and image data such as x-rays. We cannot leave the data in this form and still effectively analyze it. Rather we require an information extraction process that pulls out the required, information from the underlying sources and expresses it in a structured form suitable for analysis. Doing this correctly and completely is a continuing technical challenge.

3.3 Data Integration Aggregation and Representation

Given the heterogeneity of the flood of data, it is not enough merely to record it and throw it into a repository. Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large scale analysis all of this has to happen in a completely automated manner. This requires differences in data structure and semantics to be expressed in forms that are computer understandable and "Robotically" resolvable.

3.4 Query Processing, Data Modeling and Analysis

Methods for querying and mining Big data are fundamentally different from traditional statistical analysis on small samples. Big data is often noisy, dynamic, heterogeneous, inter-related and untrustworthy. Nevertheless, even noisy Big data could be more valuable than tiny samples because general statistics obtained from frequent patterns and correlation analysis usually overpower individual fluctuations and often disclose more reliable hidden patterns and knowledge. Further, interconnected Big data forms large heterogeneous information networks, with information redundancy can be explored to compensate for missing data, to crosscheck conflicting cases, to validate trustworthy relationships, to disclose inherent clusters and to uncover hidden relationships and models.

Mining requires integrated, cleaned, trustworthy and effectively accessible data, declarative query and mining interfaces, scalable mining algorithms and big data computing environments. At the same time, data mining itself can also be used to help improve the quality and trustworthiness of the data, understand its semantics and provide intelligent querying functions.

3.5 Interpretation

Having the ability to analyze Big data is of limited value if users cannot understand the analysis. Ultimately, a decision maker, provided the result of analysis, has to interpret these results. This interpretation cannot happen in a vacuum. Usually, it involves examining all the assumptions made and retracing the analysis. Furthermore as we saw above, there are many possible sources of error computer systems can have bugs, models almost always have assumptions, and results can be based on erroneous data. For all of these reasons, no responsible user will cede authority to the computer system. Rather he will try to understand, and verify, the results produced by the computer. The computer system must make it easy for him to do so. This is particularly a challenge with Big data due to its complexity. There are often crucial assumptions behind the data recorded. Analytical pipelines can often involve multiple steps, again with assumptions built in. The recent mortage related shock to the financial system dramatically underscored the need for such decision maker diligence rather than accept the stated solvency of a financial institution at face value, a decision maker to examine critically the many assumptions at multiple stages of analysis.

4. REMOTE SENSING BIG DATA ANALYTICS ARCHITECTURE

The term Big Data covers diverse technologies same as cloud computing. The input of Big Data comes from social networks (Facebook, Twitter, LinkedIn, etc.), Web servers, satellite imagery, sensory data, banking transactions, etc.

4.1 Remote Sensing Big Data Acquisition Unit (RSDU)

The RSDU in the remote sensing Big Data architecture that gathers the data from various satellites around the globe as shown in. It is possible that the received raw data are distorted by scattering and absorption by various atmospheric gasses and dust particles. After pre-processing phase,[9] the collected data are transmitted to a ground station using downlink channel. This transmission is directly or via relay satellite with an appropriate tracking antenna and communication link in a wireless atmosphere.

We divided the data processing procedure into two steps, such as real-time Big Data processing and off-line Big Data processing. In the case of off-line data processing, the Earth Base Station transmits the data to the data centre for storage. This data is then used for future analyses. However, in real-time data processing, the data are directly transmitted to the filtration and load balancer server (FLBS), since storing of incoming real-time data degrades the performance of real-time processing.

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 useful 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; e.g., if there is only a need for analysis of sea wave and temperature data, the measurement of these described data is filtered out, and is segmented into parts. 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

4.2 Data Processing Unit (DPU)

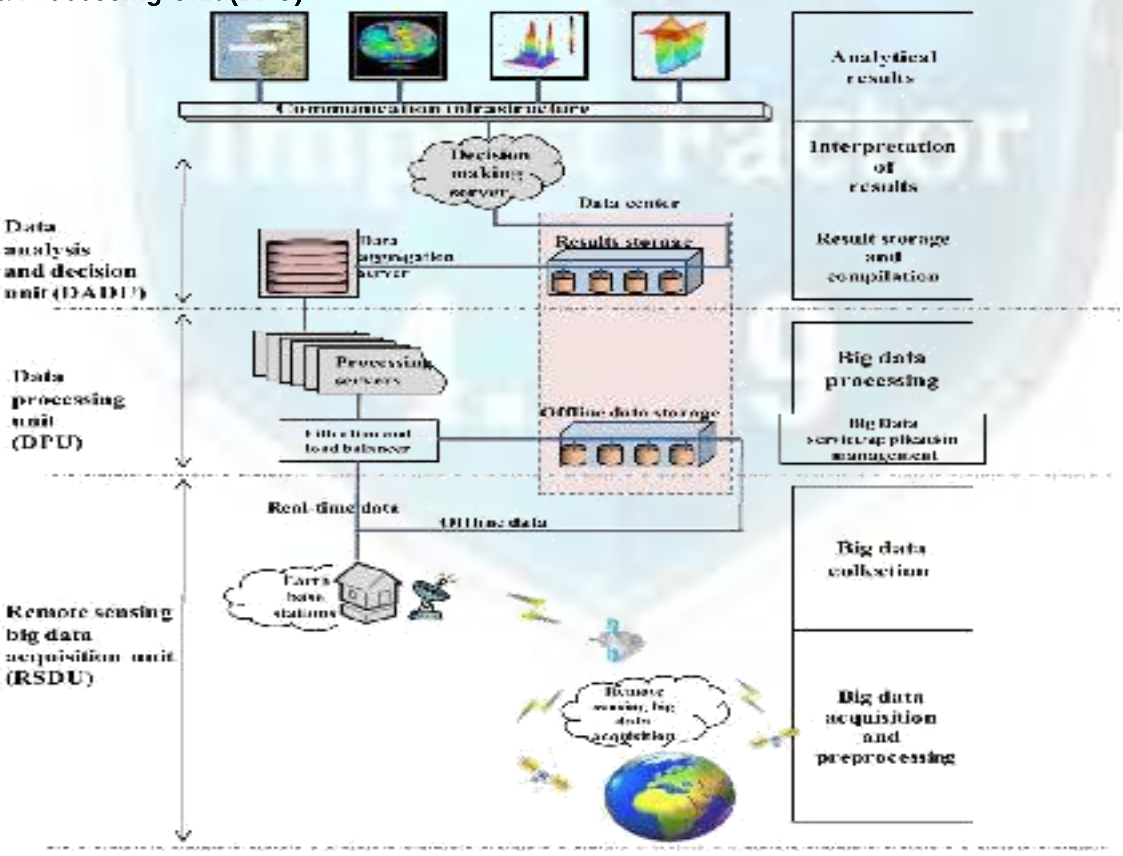

Figure:1. Remote sensing Big Data architecture.

segment of data. Since these servers perform tasks independently and in parallel, the performance proposed 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.

4.3 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. 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.

5. CONCLUSION

In this paper, we proposed architecture for real-time Big Data analysis for remote sensing application. The proposed architecture efficiently processed and analyzed real-time and offline remote sensing Big Data for decision-making. The proposed architecture is composed of three major units, such as 1) RSDU; 2) DPU; and 3) DADU. 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.

REFERENCES

[1] I. Becker-Reshef, C. Justice, "NASA Contribution to the group on earth observation (GEO) global agricultural monitoring system of systems," The Earth Obs., vol. 21, pp. 24–29, 2009.

[2] Russell G. Congalton, A review of assessing the accuracy of classifications of remotely sensed data, Remote Sens. Environ. 37 (1) (1991) 35–36.

[3] F.F. Sabins, Remote Sensing: Principles and Interpretation, Freeman, 1978.

[4] C.L. Philip Che, Chun-Yang Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, Inform. Sci. 275 (2014) 314–347.

[5] I. Gorton, Software architecture challenges for data intensive computing, in: Software Architecture, 2008. WICSA 2008. Seventh Working IEEE/IFIP Conference on, Feb. 2008, pp. 4–6.

[6] A. Labrinidis and H. V. Jagadish, "Challenges and opportunities with Big Data," in Proc. 38th Int. Conf. Very Large Data Bases Endowment, Istanbul, Turkey, Aug. 27–31, 2012, vol. 5, no. 12, pp. 2032–2033.

[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] Wikibon Blog. (Oct. 14, 2014). [2310]. Big Data Statistics [Online]. Available: wikibon.org/blog/big-data-statistics.

