IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. III (Jul.-Aug. 2017), PP 29-32
www.iosrjournals.org

Real-time Big Data Analytics and parallel processing using Hadoop on Remote Sensing data

*Ms. D.Prema swarupa rani
M. Tech Dept. of computer science and Engineering LBRCE College , Mylavaram, India
Under The Guidence of Mr. A. Raja Gopal
*Corresponding Author: *Ms. D.Prema swarupa rani*

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 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 and find Frequency occurrences*

Date of Submission: 29-06-2017 Date of acceptance: 24-07-2017

I. Introduction

A lot of enthusiasm for Big Data has risen, for the most part determined from across the board number of research issues emphatically identified with genuine applications and systems. Day by day the data is increasing very large volume from social media, videos, emails, online transitions, logs, Scientific data, mobile phones, Remote sensors and other applications. These data store in the database and grow rapidly with a massive volume of data.The great deal of interest in the field of Big Data and its analysis has risen [1]-[3]mainly driven from extensive number of research challenges strappingly related to bonafide applications[4], such as modelling , processing, querying, mining, and distributing large-scale repositories. The term "Big Data" classifies specific kinds of data sets comprising formless data, which well in data layer of technical computing applications and the Web [5]. The data stored in the underlying layer of all these technical computing application scenarios have some precise individualities in common, such as

1) large scale data
2) scalability issues
3) sustain extraction transformation loading (ETL) method and
4) development

It develop the uncomplicated interpretable analytical over Big Data warehouses with a view to deliver an intelligent and momentous knowledge for them[8]. 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 [6]-[7] These data are accumulated in databases that grow extraordinarily and become complicated to confine, form, store, manage, share, process, analyze, and visualize via typical database software tools.

II. Analyticsarchitecture

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,

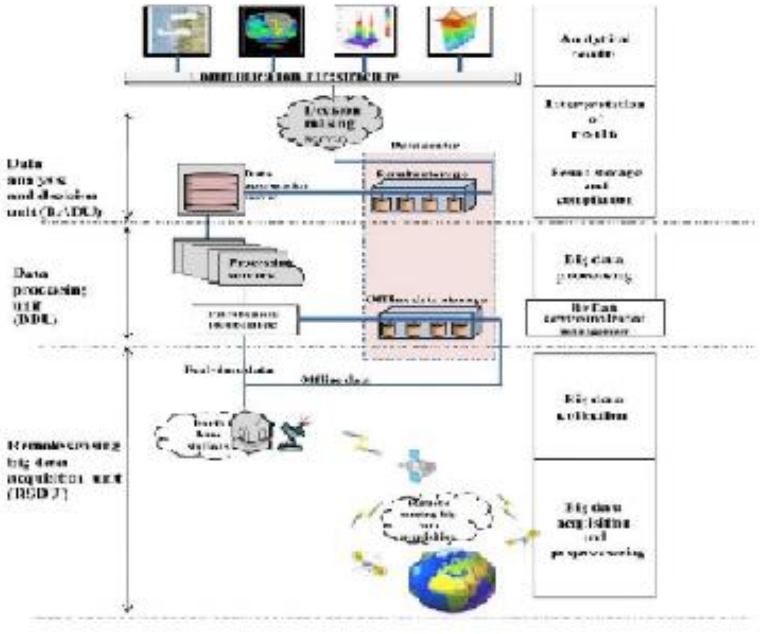

Fig: Big data analytical architecture

banking transactions, etc. Regardless of very recent emergence of Big Data architecture in scientific applications, numerous efforts toward Big Data analytics architecture can already be found in the literature. Among numerous others delineates *n* number of satellites that obtain the earth observatory Big Data images with sensors or conventional cameras through which sceneries are recorded using radiations. Special techniques are applied to process and interpret remote sensing imagery for the purpose of producing conventional maps, thematic maps, resource surveys, etc. We have divided remote sensing Big Data architecture into three parts, 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 the said parts are described as below.

A. Remote Sensing Big Data Acquisition Unit (RSDU)

Remote sensing promotes the expansion of earth observatory system as cost-effective parallel data acquisition system to satisfy specific computational requirements. The Earth and Space Science Society originally approved this solution as the standard for parallel processing in this particular context

As satellite instruments for Earth observation integrated more sophisticated qualifications for improved Big Data acquisition, soon it was recognized that traditional data processing tehnologies could not provide sufficient power for processing such kind of data. Therefore, the need for parallel processing of the massive volume of data was required, which could efficiently analyze the Big Data. For that reason, the proposed RSDU is introduced in the remote sensing Big Data architecture that gathers the data from various satellites around the globe as shown in fig. 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 make the raw data into image format, the remote sensing satellite uses Doppler or SPECAN algorithms . For effective 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 After preprocessing phase, 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. The data must be corrected in different methods to remove distortions caused due to the motion of the platform relative to the earth, platform attitude, earth curvature, non uniformity of illumination, variations in sensor characteristics, etc. The data is then transmitted to Earth Base Station for further processing using direct communication link. We divided the data processing procedure into two steps, such as real-time Big Data processing and offline Big Data processing. In the case of offline data processing, the Earth Base Station transmits the data to the data center 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. Remote sensing earth observatory image.

B. Data Processing Unit (DPU*)*

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 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.

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 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. 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 inquire Flowchart of the remote sensing Big Data architecture. 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. 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).

III. Motivation For Remote Sensing Applications

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 disea

The Data Loading and Transmission of RS "Big Data"

The massive high dimensional RS data make the data loading, memory residing as well as the data transmission among processing nodes during processing rather complicated and inefficient. Firstly, the enormous amount of RS data has far beyond the limited memory capacity of a single computer. The RS datasets are generally consist of multi-dimensional images and complex structured metadata. It is rather complicated to offer a proper and large data structure for loading and residing these massive RS data into local memory. The other problems is that the communication of the RS data blocks are common during the parallel implementing of various RS applications. Due to the limitation of the network bandwidth, the communication of RS data would be time-consuming especially when the volume of communication is extremely large.

Scheduling of large number of data-dependent tasks

For many large RS applications, like large-scale RS data mosaicking and hydrological modeling of large watershed could be described as a large collection of data dependent small tasks . The processing of these applications become extremely difficult because of the dependency among a large collection of tasks which give rise to ordering constraint. The succeeding tasks have to wait for the output data of preceding task to be available. The optimized scheduling of these bunch of tasks is critical to achieve higher performance. Therefore the problems lie in the decoupling the dependence relationships among tasks so as to achieve a minimal execution length.

Efficient and Productive Parallel Programming

The state-of-the-art cluster systems are featured with a multilevel hierarchical organization and increasing scale. Efficient and productive programming of these systems will be a challenge, especially in the context of data-intensive RS applications. Recently, several low-level parallel paradigms are extensively employed for remote sensing image processing in those multi-level hierarchical clusters. **high frequency**

IV. Conclusion And Future Detection

In 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. 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 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, we are planning to extend the proposed architecture to make it compatible 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]. D. Agrawal, S. Das, and A. E. Abbadi, “Big Data and cloud computing: Current state and future opportunities,” in Proc. Int. Conf. ExtendingDatabase Technol. (EDBT), 2011, pp. 530–533.

[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]. J. Dean and S. Ghemawat, “Mapreduce: Simplified data processing on large clusters,” Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.

[4]. H. Herodotouet al., “Starfish: A self-tuning system for Big Data analytics,” in Proc. 5th Int. Conf. Innovative Data Syst. Res. (CIDR), 2011, pp. 261–272.

[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 andStreaming Data. New York, NY, USA: Mc Graw-Hill, 2012.

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

[8]. A. Cuzzocrea, D. Saccà, and J. D. Ullman, “Big Data: A research agenda,” in Proc. Int. Database Eng. Appl. Symp. (IDEAS’13), Barcelona, Spain, Oct. 09–11, 2013.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Ms. D.Prema swarupa rani. "Real-time Big Data Analytics and parallel processing using Hadoop on Remote Sensing data." IOSR Journal of Computer Engineering (IOSR-JCE) 19.4 (2017): 29-32.