EXTENDED REAL TIME SERVICE ORIENTED SENSOR ARCHITECTURE FOR BIG DATA ANALYTICS OF SENSOR SYSTEM

J.Vinodhitha
B.E student ,
Computer Science and Engineering,
Velammal Institute of Technology
E-mail: vinodhitha@gmail.com

V.Poornima
B.E student,
Computer Science and Engineering,
Velammal Institute of Technology,
E-mail: poornimarani111@gmail.com

K.Balachander
Assistant Professor,
Computer Science and Engineering,
Velammal Institute of Technology,
E-mail: chanderbal@yahoo.com

Abstract: This paper is an attempt to enhance the existing real-time Big Data analytical architecture for remote sensing satellite application with enhancement using Social Network Data. 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 realtime, such as earthquake prediction, Tsunami prediction, fire detection, etc. Assets of real time digital world daily generate massive volume of real-time data (mainly referred to the term —Big Data||), where insight information has a potential significance if collected and aggregated effectively. In today's era, there is a great deal added to real-time re mote sensing Big Data than it seems at first, and extracting the useful inform at ion in an efficient manner leads a system toward a major The computational challenges, such as to analyze, aggregate, and store, where data are remotely collected. Keeping in vie w the above 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 propose real-time Big Data analytical architecture for processing such data environment. This paper aims to implement an intelligent architectural system to analyze and access the sensor data using Big Data analytics. As cloud resources enable the Wireless Sensor Networks to store and analyze their vast amount of data, Sensor Cloud is designed using Service Oriented Sensor Architecture. Sensor Cloud acts as an enabler for big sensor Data analytics. In the current application these three become the compelling combination. It is proposed to use the Hadoop Distributed File Systems (HDFS) concept to store the streaming sensor data on to sensor cloud for further analysis using MapReduce technique. This paper describes a public sensor cloud delivery model through cloud data analytics for sensor services. The proposed architecture acts as a Cloud Access Execution and Monitoring environment for sensor systems and is able to respond to the requested sensor client applications with greater intelligence.

KeywordHadoopDistributedFileSystems(HDF,sensorcloud,offline-dataprocessing

I. INTRODUCTION

Big data has become the new frontier of information management given the amount of data today's systems are generating and consuming. It has driven the need for technological infrastructure and tools that can capture, store, analyze and visualize vast amounts of disparate structured and unstructured data. These data are being generated at increasing volumes from data intensive technologies including, but not limited to, the use of the Internet for activities such as accesses to information, social networking, mobile computing and commerce. Corporations and governments have begun to recognize that there are unexploited opportunities to improve their enterprises that can be discovered from these data. Analytics when applied in the context of big data is the process of examining immense amounts of data, from a diverse number of data sources and in different formats, to deliver insights that can enable decisions in real or near real time. Big data analytical approaches can be employed to recognize inherent patterns, correlations and anomalies which can be discovered as a result of integrating vast amounts of data from different data sets.

Analytics when applied in the context of big data is the process of examining large amounts of data, from a variety of data sources and in different formats, to deliver insights that can enable decisions in real or near real time. Various analytical concepts such as data mining, natural language processing, artificial intelligence and predictive analytics can be employed to analyse, contextualize and visualize the data. Big data analytical approaches can be employed to recognize inherent patterns, correlations and anomalies which can be discovered as a result of integrating vast amounts of data from different data sets.

Big data analytics requires the use of new frameworks, technologies and processes to manage it. Yet its arrival in the enterprise software space has created some confusion as business leaders try to understand the differences between it and traditional data warehousing (DW) and business intelligence (BI) tools.

There are important distinctions and sufficient differentiating value between BDA and DW/BI systems which make BDA unique. Gartner defines a data warehouse as "a storage architecture designed to hold data extracted from transaction systems, operational data stores and external sources. The warehouse then combines that data in an aggregate, summary form suitable for enterprise-wide data analysis and reporting for predefined business needs."

II. RELATED WORK

Big Data and cloud computing: Current state and future opportunities[1] .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. Though scalable data management has been a vision for more than three decades and much research has focussed on large scale data management in traditional enterprise setting, cloud computing brings its own set of novel challenges that must be addressed to ensure the success of data management solutions in the cloud environment. This tutorial presents an organized picture of the challenges faced by application developers and DBMS designers in developing and deploying internet scale applications. Our background study encompasses both classes of systems: (*i*) for supporting update heavy applications, and (*ii*) for ad-hoc analytics and decision support. We then focus on providing an in-depth analysis of systems for supporting update intensive web-applications and provide a survey of the state-of-the-art in this domain. We crystallize the design choices made by some successful systems large scale database management systems, analyze the application demands and access patterns, and enumerate the desiderata for a cloud-bound DBMS.
Mad skills: New analysis practices for Big Data[2]. As massive data acquisition and storage becomes increasingly affordable, a wide variety of enterprises are employing statisticians to engage in sophisticated data analysis. In this paper we highlight the emerging practice of Magnetic, Agile, Deep (MAD) data analysis as a radical departure from traditional Enterprise Data Warehouses and Business Intelligence. We present our design philosophy, techniques and experience providing MAD analytics for one of the world's largest advertising networks at Fox Audience Network, using the Greenplum parallel database system. We describe database design methodologies that support the agile working style of analysts in these settings. We present dataparallel algorithms for sophisticated statistical techniques, with a focus on *density* methods. Finally, we reflect on database system features that enable agile design and flexible algorithm development using both SQL and MapReduce interfaces over a variety of storage mechanisms.
Mapreduce: Simplified data processing on large clusters [3].MapReduce is a programming model and an associated implementation for processing and generating large datasets that is amenable to a broad variety of real-world tasks. Users specify the computation in terms of a *map* and a *reduce* function, and the underlying runtime system automatically parallelizes the computation across large-scale clusters of machines, handles machine failures, and schedules inter-machine communication to make efficient use of the network and disks. Programmers find the system easy to use: more than ten thousand distinct MapReduce programs have been implemented internally at Google over the past four years, and an average of one hundred thousand MapReduce jobs are executed on Google's clusters every day, processing a total of more than twenty petabytes of data per day.
Starfish: A self-tuning system for Big Data analytics[4].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. Most practitioners of big data analytics—like computational scientists, systems researchers, and business analysts—lack the expertise to tune the system to get good performance. Unfortunately, Hadoop's performance out of the box leaves much to be desired, leading to suboptimal use of resources, time, and money (in payas-you-go clouds). We introduce Starfish, a self-tuning system for big data analytics. Starfish builds on Hadoop while adapting to user needs and system workloads to provide good performance automatically, without any need for users to understand and manipulate the many tuning knobs in Hadoop. While Starfish's system architecture is guided by work on self-tuning database systems, we discuss how new analysis practices over big data pose new challenges; leading us to different design choices in Starfish.
Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming[5]. Big Data represents a new era in data exploration and utilization, and IBM is uniquely positioned to help clients navigate this transformation. This book reveals how IBM is leveraging open source Big Data technology, infused with IBM technologies, to deliver a robust, secure, highly available, enterprise-class Big Data platform. The three defining characteristics of Big Data--volume, variety, and velocity--are discussed. You'll get a primer on Hadoop and how IBM is hardening it for the enterprise, and learn when to leverage IBM InfoSphere BigInsights (Big Data at rest) and IBM InfoSphere Streams (Big Data in motion) technologies. Industry use cases are also included in this practical guide. Learn how IBM hardens Hadoop for enterprise-class scalability and reliability Gain insight into IBM's unique in-motion and at-rest Big Data analytics platform Learn tips and tricks for Big Data use cases and solutions Get a quick Hadoop primer.

III. EXISTING SYSTEM

Most recently designed sensors used in the earth and planetary observatory system are generating continuous stream of data. Moreover, majority of work have been done in the various fields of remote sensory satellite image data, such as change detection, gradient-based edge detection, region similarity based edge detection, and intensity gradient technique for efficient intraprediction.

IV. PROPOSED SYSTEM

In this paper, we referred the high speed continuous stream of data or high volume offline data to "Big Data," which is leading us to a new world of challenges. This paper presents a remote sensing Big Data analytical architecture, which 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 useful information is transmitted to the Earth Base Station for further data processing. Earth Base Station 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 transmitted to the filtration and load balancer server, where filtration algorithm is employed, which extracts the useful 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 .The proposed architecture and the algorithms are implemented in Hadoop using MapReduce programming by applying remote sensing earth observatory data. 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.

A. REMOTE SENSING BIG DATA ACQUISITION

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 technologies could not provide sufficient power for processing such kind of data. The need for parallel processing of the massive volume of data was required, which could efficiently analyze the Big Data.

B .DATA PROCESSING

In data processing unit (DPU), the filtration and the 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 assigns them to various processing servers.

C. DATA ANALYSIS AND DECISION

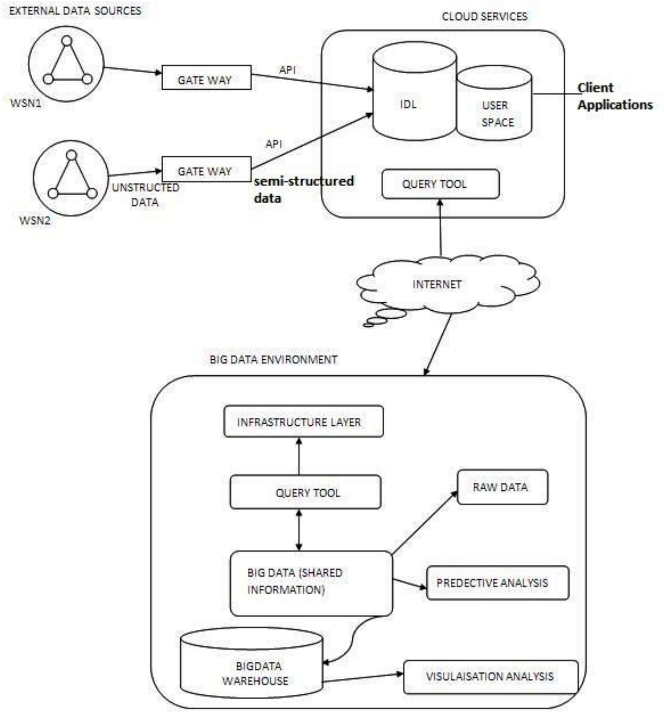
DADU contains three major portions, such as aggregation and compilation server, results storage server(s), and decision making server. When the results are ready for compilation, the processing servers in DPU send the partial results to the aggregation and compilation server, since the aggregated results 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.

D.PERFORMANCE ANALYSIS

In the last module, we implement performance Analysis with Intelligent Graph component, showing the map-reduce data in graphical output with Performance Analysis on various graphs.

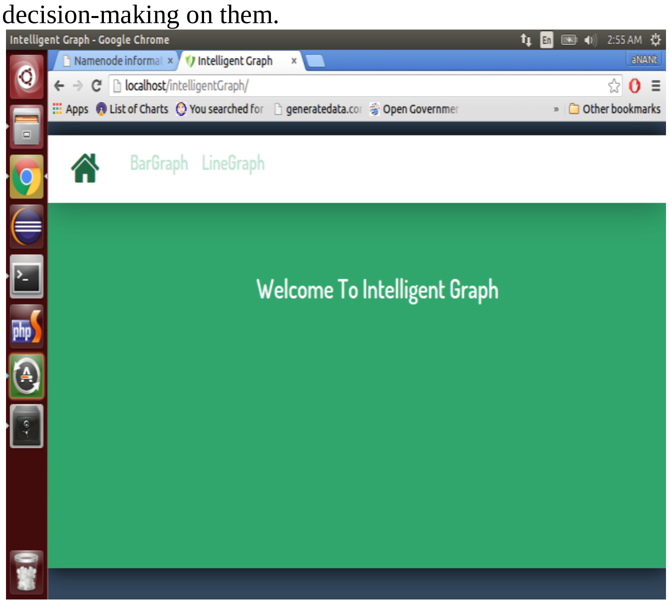
V. ARCHITECTURE

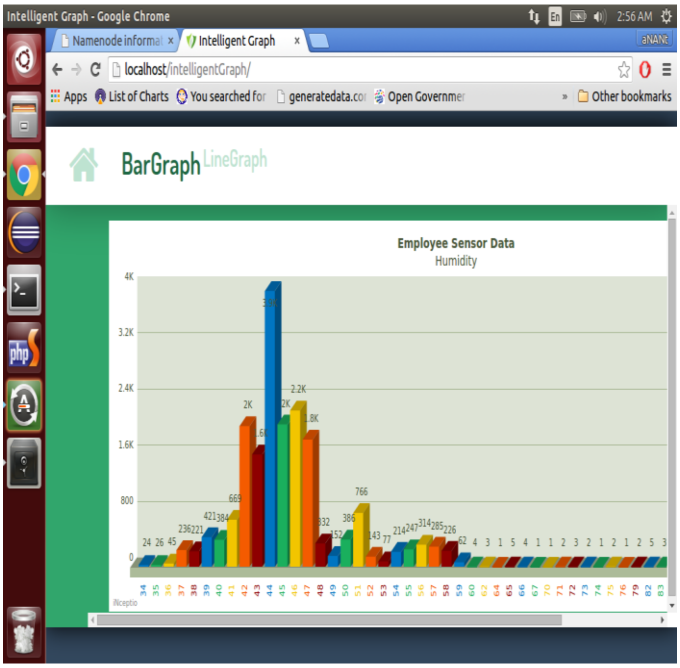
At first, the data are remotely pre-processed, which is then readable by the machines. Afterward, this useful information is transmitted to the Earth Base Station for further data processing. Earth Base Station performs two types of processing, such as processing of real-time and offline data. In case 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 transmitted to the filtration and load balancer server, where filtration algorithm is employed, which extracts the useful 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 proposed architecture and the algorithms are implemented in Hadoop using MapReduce programming by applying remote sensing earth observatory data.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.

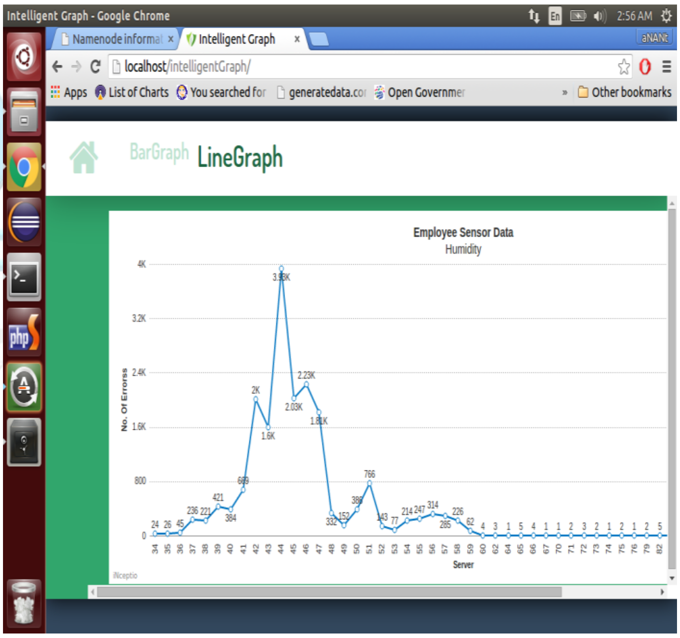


VI.RESULTS AND IMPLEMENTATION

We implemented our algorithms in simple java language using Beam-5.0 library as well in Hadoop using MapReduce, initially in a single-node environment. In the Hadoop implementation, Map function takes the image block offset as a key and the image block (pixel values) as a value parameter. Since Hadoop MapReduce cannot directly process image blocks, the whole product image data are converted into sequence file to be processed using MapReduce. In such a way,one line of the sequence file contains one image block. Map function performs parameters calculations on incoming block values and finally sends the block number as a key and list of parameters results as a value to the Reduce function. Reduce function uses parameter results for performing decision-making on them.







VIII. 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. 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.

IX. REFERENCES

[1] 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.

[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. Herodotou *et 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] 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.

[6]Real-Time Big Data Analytical Architecture for Remote Sensing Application Muhammad Mazhar Ullah Rathore, Anand Paul, *Senior Member, IEEE*, Awais Ahmad, *Student Member, IEEE*, Bo-Wei Chen, *Member, IEEE*, Bormin Huang, and Wen Ji, *Member, IEEE*

[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.

[9] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 2nd ed. New York, NY, USA: Academic Press, 1997.

[10] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003.

[11] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[12] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer, 2006.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, 2009.

[18] S. Kalluri, Z. Zhang, J. JaJa, S. Liang, and J. Townshend, "Characterizing land surface anisotropy from AVHRR data at a global scale using high performance computing," *Int. J. Remote Sens.*, vol. 22, pp. 2171–2191, 2001.