Volume 6, Issue 12, December 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Effective Error Detection System

Vishakha Bajad.
M. E. Second Year Student
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, Maharashtra, India

Jayshree R.Pansare.
Assistant Professor
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, Maharashtra, India

Abstract—Big sensor data is increasing hope that analyses of this massive amount of data will provide insight that is valuable to business, society and scientific research. In Big data it is not possible to manage whole database once using traditional database management system. Cloud computing is part of computer science and it enable providing Internet services to external customer via very scalable computing capacities. Some technique has been developed in recent years for processing sensor data on cloud, such as sensor cloud. But, those strategies do no longer offer guide on fast detection and locating of errors in large sensor records. For quicker error detection in large sensor facts units, on this system, it develop a novel data errors detection approach which offers the total feature of cloud platform and the community feature of Wi-Fi or wireless sensor network (WSN). Firstly different types of errors classification is conducted. The error detection is mainly based on scale free network topology and most of the detection operation is done using clustering method. Hence the detection and location process can be dramatically accelerated. After finding the errors data cleaning and recovery is done. For the experiential purpose U-cloud is used, it is demonstrated that the proposed approach can significantly reduce the time for error detection and locating in big data set with acceptable error detection accuracy.

Keywords -Big data, scalable computing, time efficiency, sensor data and complex network systems.

I. INTRODUCTION

The term big data refers to large amount of information which is collected by us and our surroundings. Big data varies depending on the capabilities of the user and their tools.2.5 quintillion bytes of data were created every day, and 90% of the data is created in the last two years alone [1],[6]. Big data requires a set of techniques and technologies with new forms of integration. The latency requirement is the main consideration for big data platform. data errors is a challenging issue. Some work for big data analysis and error detection in complex networks including intelligence sensors networks has been done [7], [8]. Error detection and debugging with online data processing techniques has been also done [9], [10]. Since these techniques not deal with big data, so they were unable to cope with current dramatic

Increase of data size. Some work has been done about processing sensor data on cloud [3], [8]. However, fast detection of errors in big data remains challenging. The error detection approach in this paper will be based on neural networks. This paper is organized as follows. Related work of existing system is present in the Section II. Outlines the proposed system and the workflow of the proposed work process along with algorithm are included in section III. Presents the results to validate the performance. Finally, some brief conclusions.

II. RELATED WORK

In our proposed approach, the error detection is based on the scale-free network topology and most of detection operations can be conducted in limited temporal or spatial data blocks instead of a whole big data set. Hence the detection and location process can be dramatically accelerated. Furthermore, the detection and location tasks can be distributed to cloud platform to fully exploit the computation power and massive storage. Through the experiment on our cloud computing platform of U- Cloud, it is demonstrated that our proposed approach can significantly reduce the time for error detection and location in big data sets generated by large scale sensor network systems with acceptable error detecting accuracy.

A. Big data set processing on cloud

One essential quality of cloud computing is in aggregation of resources and data into data centres on the Internet. The present cloud[1][10] services (IaaS, PaaS and SaaS) realize improved execution efficiency by aggregating application execution environments at various levels including server, OS and middleware levels for sharing them. Meanwhile, another approach of aggregating data into clouds has also been launched, and it is to analyse such data with the powerful computational capacity of clouds

With the rapid growth of emerging applications like social network analysis, semantic Web analysis and bioinformatics network analysis, a variety of data to be processed continues to witness a quick increase. Effective management and analysis of large-scale data poses an interesting but critical challenge. Recently, big data has attracted a lot of attention from academia, industry as well as government. This paper introduces several [2][5] big data processing techniques from system and application aspects. First, from the view of cloud data management and big data processing mechanisms, we present the key issues of big data processing, including cloud computing platform, cloud architecture, cloud database and data storage scheme. Following the MapReduce [2] parallel processing framework, we then introduce MapReduce optimization strategies and applications reported in the literature. Finally, we discuss the open issues and challenges, and deeply explore the research directions in the future on big data processing in cloud computing environments.

B. Analyze Error in Sensor Networks and Complex Networks

As an imperative logical enormous information source, scientific sensor systems and wireless sensor network applications produce a variety of enormous data sets in real time through various monitored activities in different application domains, such as healthcare, military, environment, and manufacturing. With the sensational increment of huge information produced from complex system frameworks, to find and locate the errors in big data sets turns to be difficult with normal computing and network systems. Wang et al. [2] provide a classification for errors on social networks based on error scenarios analysis. Hence, the error models and types presented in [2] can be extended for the errors in complex network systems. Xiong proposed an approach which can be used to detect the text data errors in data sets of social network. Mukhopadhyay proposed a model based error correction method for WSN. It is conducted over intelligent sensor network itself. This technique is based on the correction with data trend prediction. The primary goal of this location error analysis is to demonstrate the practical use of the location errors for optimal resource consumption. It can be concluded that current data error detection techniques for complex network systems focus on in-network detecting with intelligent nodes or offline analysis at the root. They ignore the scalability, massive resource and powerful computation capability provided by Hadoop. The proposed approach in this paper aims to address this issue by utilizing the inherent features of Map-Reduce.

III. PROPOSED SYSTEM

A. System Architecture

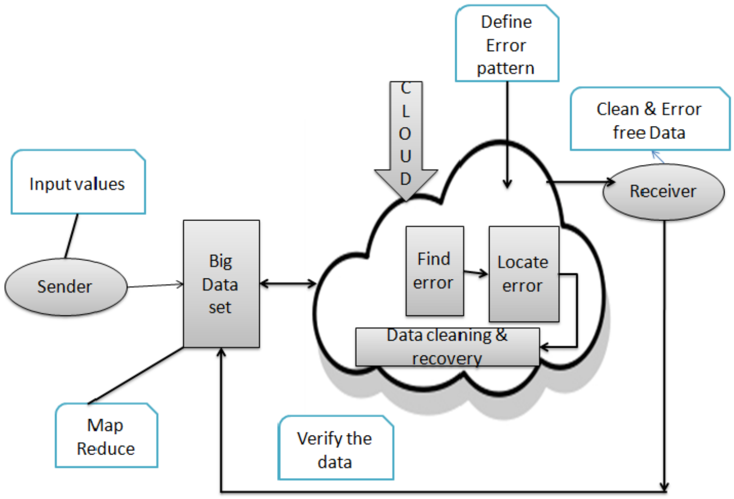


Figure 1: Effective Error Detection System

- In fig. there is a complex network and cloud platform for running error detecting algorithms.
- Input data is collected from the initial node i.e. sender the data uploaded on a cloud platform. Data must be in numeric form.
- In big data set it contains a variety of data coming from a various sources. MapReduce algorithm is applied on clusters which are made up o big data.
- Different operations are performed on the cloud platform such as, error detection, location finding, data cleaning and error recovery.

During the filtering of big data sets, whenever an anomalous data is encountered, detection algorithm has to perform two tasks. They are illustrated as two functions here. "fd(n/e,t) " is decision making function which determines whether detected anomalous data is a true error. In other words, fd(n/e,t) has two outputs, "false negative" for detecting a true error and "false positive" for selecting a non-error data. "fl (n/e,t)" is a function for tracking and returning original error source. As above two functions results with error detection the process gets successfully completed. In this paper the sensor data values are collected and processed, where we do error detection. For detecting errors, we approach for spatial and temporal correlation models .The error correction will be the future work. It shows a sign of fast Error Recovery as modern technique Map reduce is adopted The proposed error detection approach in this paper will be based on the classification of error types. Specifically numerical data errors are set down and introduced in big data error detection approach. The defined error model will trigger the error detection process, compares to previous error detection of sensor network systems. It will be designed and developed by utilizing the massive data processing capability. In addition, the architecture feature of complex networks will also be analysed to combine with the parallel computing with a more efficient way.

B. Algorithm

i. Error Detection Using Neural Network:

Neural network is used to train the data which is received from the sensors and then it detects incorrect information. It increase's its reliability and updates the training data. Error detection with new data an d neural network's result is used to improve the accurate data collection.

ii. Error Localization:

After error detection, it is important to locate the position and source of the detected error in the original WSN graph G(V, E). The inputs are the original graph of a scale-free network G (V, E), and an error data D. The output is G'(V', E') which is thesubset of the G to indicate the error location and source.

iii. Error Correction:

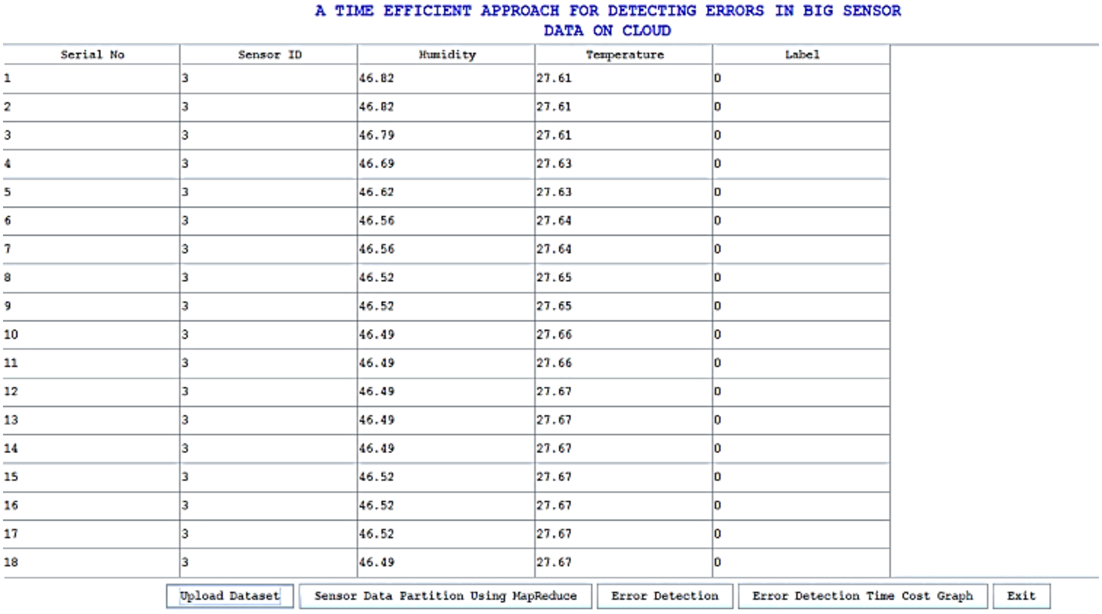
Forward Error Correction (FEC) codes can detect and correct a limited number of errors without a feedback channel. Block codes and Convolution codes are its types. BCH code is the significant example of it. For the purpose of detecting and checking the errors, BCH encodes k data bits into n code bits by adding n-k parity checking bits.

Given the length of the codes is n = 2m-1 for any integer m=3, we will have t (where t<2m-1), is the bound of the error correction. BCH can correct any combination of errors (burst or separate) fewer than t in the n-bit-codes. The number of parity checking bits is n-k = mt.

iv. Data Recovery:

After finding all the error type's different operation is conducted over it to get the error free data. All the corrupted data is recovered.

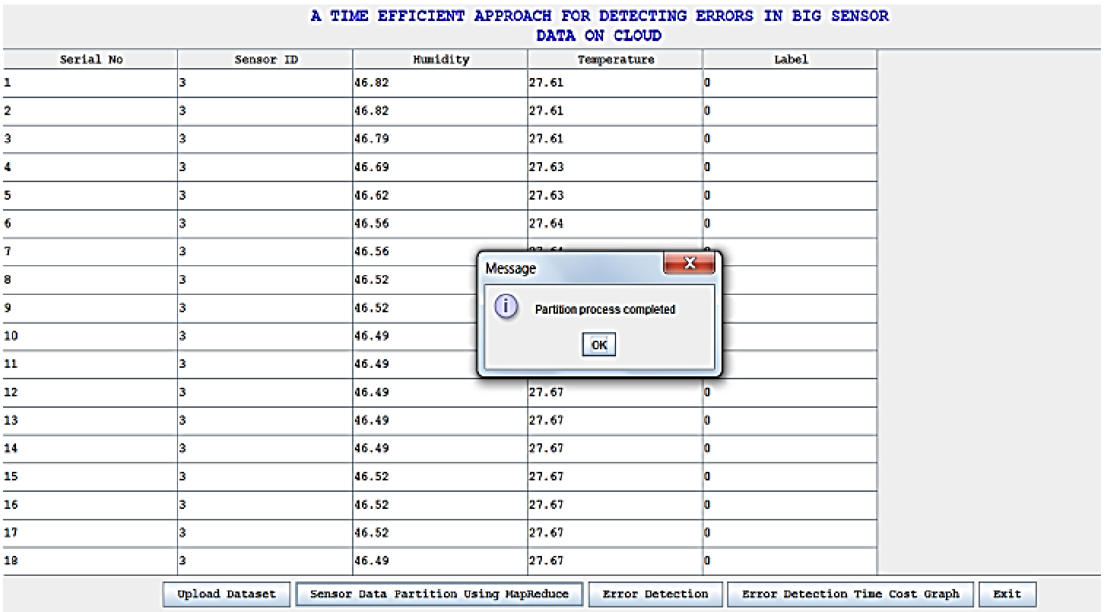
C. Expected Result



A TIME EFFICIENT APPROACH FOR DETECTING ERRORS IN BIG SENSOR DATA ON CLOUD

Serial No	Sensor ID	Humidity	Temperature	Label
1	3	46.82	27.61	0
2	3	46.82	27.61	0
3	3	46.79	27.61	0
4	3	46.69	27.63	0
5	3	46.62	27.63	0
6	3	46.56	27.64	0
7	3	46.56	27.64	0
8	3	46.52	27.65	0
9	3	46.52	27.65	0
10	3	46.49	27.66	0
11	3	46.49	27.66	0
12	3	46.49	27.67	0
13	3	46.49	27.67	0
14	3	46.49	27.67	0
15	3	46.52	27.67	0
16	3	46.52	27.67	0
17	3	46.52	27.67	0
18	3	46.49	27.67	0

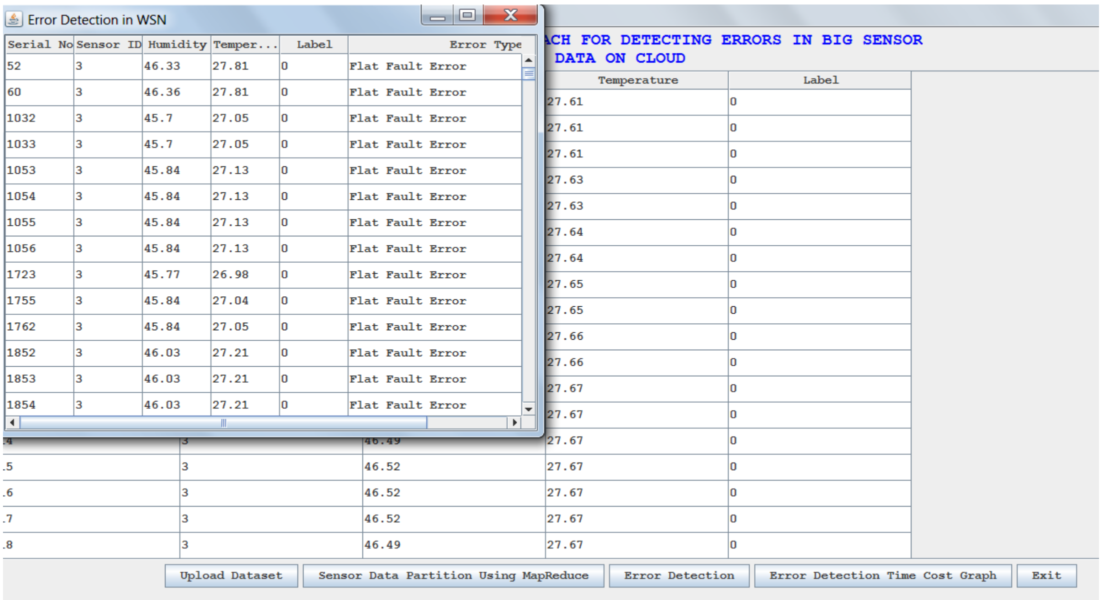
Data view on Cloud



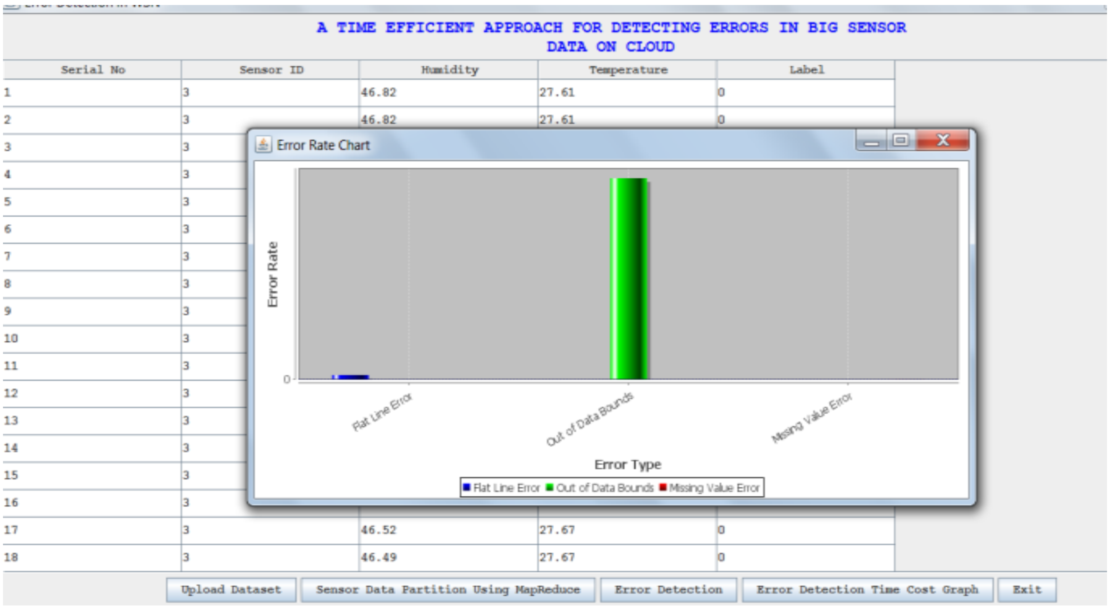
A TIME EFFICIENT APPROACH FOR DETECTING ERRORS IN BIG SENSOR DATA ON CLOUD

Serial No	Sensor ID	Humidity	Temperature	Label
1	3	46.82	27.61	0
2	3	46.82	27.61	0
3	3	46.79	27.61	0
4	3	46.69	27.63	0
5	3	46.62	27.63	0
6	3	46.56	27.64	0
7	3	46.56	[illegible]	[illegible]
8	3	46.52	[illegible]	[illegible]
9	3	46.52	[illegible]	[illegible]
10	3	46.49	[illegible]	[illegible]
11	3	46.49	[illegible]	[illegible]
12	3	46.49	27.67	0
13	3	46.49	27.67	0
14	3	46.49	27.67	0
15	3	46.52	27.67	0
16	3	46.52	27.67	0
17	3	46.52	27.67	0
18	3	46.49	27.67	0

Partition Process Completed



Error Detection Data



Graph showing various Error Stages

IV. CONCLUSION

In order to detect and find the location of error in big data set mainly uses a sensor network systems, a novel approach is developed with cloud computing. Firstly the classification of error in big data sets is presented. Secondly, the correlation and comparison between sensor network systems and the scale-free complex networks are introduced. According to each define error type and the features from scale-free networks, the system proposed different error detection strategies for detecting and locating errors in big data sets on cloud. All the process of error detection is conducted over the User defined cloud i.e. U-cloud. The significance for the system is: 1It gives the fast error detection and locating using time efficient approach and 2) after error detection process done effectively recover the corrupted file to avoid loss of data.

REFERENCE

[1] D.J. Wang, X. Shi, D.A. Mcfarland, and J. Leskovec, "Measuremen Error in Network Data: A Re-Classification," *Social Networks*, vol. 34, no. 4, pp. 396-409, Oct. 2012.

[2] S. Slijepcevic, S. Megerian, and M. Pot.konjak, "Charaterization of Lacation Error in Wireless Sensor Networks: Analysis and Application," *Proc. the Second Int'l Conf. Information Processing in Sensor Networks* (*IPSN'03*), pp. 593-608, 2003.

[3] A. Sheth, C. Hartung, and Richard Han, "A Decentralized Fault Diagnosis System for Wireless Sensor Netowrks," *Proc. IEESecond Conf. Mobile Ad-hoc and Sensor Systems (MASS '05),* Nov. 2005.

[4] M.M.H. Khan, H.H.K. Le, H. Ahmadi, T.F. Abdelzaher, and J. Han, "Dustminer: Troubleshooting Interactive Complexity Bugs in Sensor Networks," *Proc. ACM Sixth Conf. Embedded Network Sensor Systems (SenSys '08),* pp. 99-112, 2008.

[5] X. Zhang, T. Yang, C. Liu, and J. Chen, "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using Systems, in MapReduc on Cloud," *IEEE Trans. Parallel and Distributed*, vol. 25, no. 2, pp. 363-373, Feb. 2014

[6] N. Laptev, K. Zeng, and C. Zaniolo, "Very Fast Estimation for Result and Accuracy of Big Data Analytics: The EARL System," *Proc. IEEE 29th Int'l Conf. Data Eng. (ICDE),* pp. 1296-1299, 2013.

[7] K.H. Lee, Y.J. Lee, H. Choi, Y.D. Chung, and B. Moon, "Parallel Data Processing with MapReduce: A Survey," *ACM SIGMOD Record,* vol. 40, no. 4, pp. 11-20, 2012.

[8] X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen, "A Privacy Leakage Upper-Bound Constraint Based Approach for Cost-effective Privacy Preserving of Intermediate Datasets in Cloud," *IEEE Trans. Parallel and Distributed Systems,* vol. 24, no. 6, pp. 1192-1202,June201

[9] K. Shim, "Map Reduce Algorithms for Big Data Analysis," Proc. VLDB Endowment, vol. 5,no. 12, pp. 2016-2017, 2012

[10] J.R.Pansare, V.D. Bajad "Error Detection in Big Sensor Data on Cloud using Time Efficient Technique*" Paper Presented In Research Advanced In Computer Engineering* at MESCOE,Pune. January,2016

[11] J.R. Pansare,V.D.Bajad "Effective Error Detection on Cloud Using Time Efficient Technique" *Paper Published n International Conference ACM Women In Research(ACM WIR), ACM Digital Library* ISBN: 978-1-4503-4278-0 Pages12-14,March, 2016

[12] V. D. Bajad, J. R. Pansare, "Errors Detection in Big Sensor Data on Cloud using Time Efficient Technique", *Paper Presented in cPGCON at PCCOE,* Pune, March, 2016.

[13] V. D. Bajad," Effectively Error Detection on Cloud by using Time Efficient Technique" *Paper published International Journal of Computer Science Trends and Technology (IJCST),* ISSN: 2347-8578, Volume 4 Issue 5, page 30, Sep - Oct 2016.

