The Data Analysis and Encrypting Data using Efficient Algorithm in Big Sensor Network Data

[1]B. Bhavani, [2]V. Mahesh, [3]Dr. A. Chandra Sekhar
[1,2,3]Dept of CSE, Avanthi Institute of Engineering and Technology, Vizianagaram, A.P, India

Abstract
Detecting error in the substantial volume of data is the most complex process where the quantity of data's are develops in size. In the current work, time proficient approach is proposed to detect the errors dwells in the big sensor data where the gathered data would be apportioned into various parcel and the errors will be recognized and situated by contrasting it and the error designs which are predefined. However this work couldn't right the errors which obliges sender to retransmit the data again which may expand the time many-sided quality. This issue is overcome in the proposed philosophy by presenting the forward error rectification strategy which will redress the errors exhibit in the big sensor data's naturally. In the proposed framework, we propose neural system calculation for error detection powerfully instead of existing framework. The error detection is utilized to decrease the mistaken data by blame sensors in big data set. The proposed approach is expanding the effectiveness and unwavering quality of big sensor data. From the exploratory outcome, the conclusion says that the proposed framework is better than existing framework by method for higher execution. In this paper, we build up a novel data error detection approach which abuses the full calculation capability of cloud stage and the system highlight of WSN. Firstly, an arrangement of sensor data error sorts are ordered and characterized. In light of that grouping, the system highlight of a bunched WSN is acquainted and examined with bolster quick error detection and area. In particular, in our proposed approach, the error detection depends on the without scale arrange topology and the majority of detection operations can be led in constrained transient or spatial data hinders rather than an entire big data.

Keywords
Big data, Cloud Computing, Sensor Networks, Data Abnormality, Error Detection.

I. Introduction
An aggregation of data sets so tremendous and composite that it discovers the opportunity to be handle with standard data prepare and association applications in big data. Big data addresses the progress of the human quick limit, procedure to get, immediate, and handle the data inside a snuck past time. Handling standards on big data at present contrast at the basic level of reflection on whether the gathering mode, or logically/close consistent on spouting (data that is industriously coming in and should be dealt with rapidly). Around there, we incorporate two particular foundations: Hadoop for error dealing with and Ubuntu for relentless prepare. MapReduce is a programming model and a related execution for dealing with and making boundless datasets [1]. Big data contrasts from customary data in different estimations: (i) Quantity of data sources (ii) Heterogeneous nature of data sources (iii) Dynamic nature of data sources that is upgrading quickly (iv) characteristics of data sources moves in different focuses. Circulated processing gives a best stage to arranging data which is intricate. Impermanent utilize and limit on interest are fundamental properties of cloud which makes it convincing for arranging big data. For prepare big data applications, security is essential which is given utilizing cloud [2]. Big data constructs and strategize to keep advancing at a quick pace, yet the significant progressions they depend upon have, generally speaking, been imagined different years before. The amazingly augmented digitization of human improvement and machine-to-machine trades, joined with liberal scale mild equipment, is making commonsense different successfully scholastic thoughts of parallel and passed on get ready, close by new changes basic to make them on a very basic level additionally obliging in veritable applications [3]. The exactness of a classifier on a given test set is the rate of tuples that are asked for definitely (these test set tuples with names ought not to have been utilized to set up the classifier). So also, the exactness of a pointer suggests how well a given marker can figure the estimation of the anticipated trademark for new or up to this time unnoticeable data. The error rate or misclassification rate of a classifier is essentially whatever remains of the rate of tuples that were not grouped accurately [4]. Distributed computing foundation is transforming into across the board as a consequence of it gives an open, adaptable, ascendible and reconfigurable stage. The anticipated error detection approach amid this paper will be upheld the grouping of error sorts. In particular, nine sorts of numerical data irregularities/errors are recorded and presented in our cloud error detection approach. The sketched out error model can trigger the error detection strategy. Contrasted with past error detection of sensor system frameworks, our approach on cloud will be composed and created by using the huge preparing ability of cloud to support error detection speed and continuous response. Furthermore, the outline highlight of entangled systems likewise will be broke down to blend with the distributed computing with an extra temperate strategy. In view of current examination writing survey, we tend to partition confused system frameworks into without scale sort and non sans scale kind. Sensor system could be a sensibly without scale convoluted system framework that matches cloud quantifiability include. Our arranged error detection approach on cloud is particularly cut for discovering errors in enormous data sets of sensor systems. The primary commitment of our arranged detection is to achieve huge time execution change in error detection while not trading off error detection precision.

II. Related Work
For Analyzing shows of models from various edges, related organization for recognizable proof of error, big data anticipating cloud, for complex system structures will be explored and contemplated. a) Big data handling A procedure of keeping up what you look like at any issue in these works is their versatility to an extensive measure of data. Computations have extended their various quality to beat more personality boggling techniques. This makes level of counts obliged to log off disclosure. The Big Data fundamentals are found not just in mammoth associations, for example, Amazon or Google, yet in different insignificant business attempts that require tending to, stockpiling and recovery over extensive scale structures. It is at present fundamental to see the computation in parallel; utilizing contemplations, for example,

MapReduce [5] for better change. Circulated processing gives a faultless stage to inducing of big data, stockpiling and unwinding with its colossal estimation control [6]. It is unavoidable to experience the issue of managing big data in different veritable applications. These days remarkable sort of work has been capable for arranging big data with cloud. A normal cloud based appropriated structure for big data dealing with is Amazon EC2 base as an association. A scattered stockpiling is upheld by Amazon S3. MapReduce [7] is clutched as a programming model for big data dealing with over conveyed processing. The issue of dealing with incremental big data is researched at different focuses from different points of view. b) WSN handling in connection with cloud At the moment that data from liberal sensor systems is should have been gathered and observed remotely sensor-Cloud is beneficial for a couple of uses. For ecological checking, social insurance, business exchanges, transportation, WSN connects with innovative blueprints. Remote sensor system structures have created assorted courses of action in various fields, for example, cataclysm watching, catastrophe warming, natural studying, and business change strategy and data gathering. Sensor cloud masterminds has been conveyed to set up the remote sensor data gathered by WSN. Plan of sensor cloud is helpful in different applications for the most part when the data is found remotely. Big data is hard to get ready utilizing close to database association devices since volume of big data is developing quickly with accumulation in data sets [8]. c) Error detection in systems Data error is unavoidable in different certifiable complex structure structures. To discover and find errors in big data sets winds up being incredibly taking a stab at undertaking with typical computational forces of standard structures as there is energetic improvement of big data conveyed from complex system structures, for example, interpersonal affiliations and huge scale sensor structures. Wang et al. give an essential gathering to errors on interpersonal relationship in context of error conditions examination which traces the lead of error conditions. This bunching combines 6 sorts of standard errors with missing data or errordata. Nature of four focus level system measures is looked this grouping structure [9]. Mukhopadhyay [10] proposed a model based error change method for Wireless sensor system. Shrewd sensor systems are utilized as a bit of this rectification technique. This structure depends on upon the change with data configuration evaluate. To locate the essential driver of errors is as fundamental as perceiving and curing error. To separate concealed driver of error, an instrument a sensor system looking at is utilized. Notwithstanding, the things which should be enhanced are client interface, adaptability and time execution.

III. Methodology

A. Big Data Processing on Cloud

With the fast development of modern information technology, we enter a new era of data. Hence, the technique to process big data has become a fundamental and critical challenge for modern society. Cloud computing can be regarded as an ingenious combination of a series of developed or developing ideas and technologies, establishing a pay-as-you-go business model by offering IT services using economies of scale. Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing provides an ideal platform for big data storage, dissemination and interpreting with its massive computation power .In many today"s real world applications, such as social networks, complex network monitoring, the scientific analysis of protein interactions and wireless sensor networks self monitoring, it is unavoidable to encounter the problem of dealing with big data and big data streams on cloud. At present, some work has been done for processing big data with cloud. Amazon EC2infrastructure as a service is a typical cloud based distributed system for big data processing. Amazon S3 supports distributed storage. Map Reduce is adopted as a programming model for big dataprocessing over cloud computing. Plenty of recent research has investigated the issues of processing incremental data on cloud. Kienzler et al. [8] designed a "stream-as-you-go" approach to access and process on incremental data for dataintensive cloud applications via a stream-based data management architecture. The extension of the traditional Hadoop framework [3] to develop a novel framework named Incoop by incorporating several techniques like task partition and memorization-aware schedule. Olston et al. [9] present a continuous workflow system called Nova on top of Pig/Hadoop through state full incremental data processing. Map Reduce has been widely revised from a batch processing framework into a more incremental one to analyze hugevolume of incremental data on cloud. It is a framework for processing parallelizable problems across big data sets using a large number of computers (nodes), collectively referred to as a cluster in which all computers (nodes) are on the same local network

B. On-Cloud Processing for WSN

Recently, wireless sensor network systems have been used in different areas, such as environment monitoring, military, disaster warning and scientific data collection. In order to process the remote sensor data collected by WSN, sensorcloud platform has been developed including its definition, architecture, and applications. Due to the features of high variety, volume, and velocity, big data is difficult to process using on- hand database management tools or traditional sensor-cloud platform. Big data sets can come from complex net- work systems, such as social network and large scale sensor networks. In addition, under the theme of complex network systems, it may be difficult to develop time- efficient detecting or trouble-shooting methods for errors in big data sets, hence to debug the complex network systems in real time. Sensor-Cloud [6] is a unique sensor data storage, visualization and remote management platform that leverages powerful cloud computing technologies to provide excellent data scalability, fast visualization, and user programmable analysis. Initially, sensor-cloud was designed to support longterm deployments of Micro- Strain wireless sensors. But nowadays, sensor-cloud has been developed to support any web-connected third party device, sensor, or sensor network through a simple Open Data API. Sensor-Cloud can be useful for a variety of applications, particularly where data from large sensor networks needs to be collected, viewed, and monitored remotely. For example, structural health monitoring and condition-based monitoring of high value assets are applications where commonly available data tools often come up short in terms of accessibility, data scalability, programmability, or performance. Sensor-Cloud represents a direction for processing and analyzing big sensor data using cloud platform. The online WSN data quality and data cleaning issues are discussed in [1-2] by Elnahrawy and path. They deal with the problems of outliers, missing information, and noise. A novel online approach for modeling and online learning of temporal-spatial data correlations in sensor networks is developed. A Bayesian approach for reducing the effect of noise on sensor data online is also pro- posed [5]. The proposed approach is efficient in reducing the uncertainty associated with noisy sensors. However, the scalability and error detection accuracy are not dealt.

C. Data Error Detection in Sensor Networks and Complex Networks

As an important scientific big data source, scientific sensor systems and wireless sensor network applications produce a variety of large data sets in real time through various monitored activities in different application domains, such as healthcare, military, environment, and manufacturing. In many real world complex network systems, data error is unavoidable. With the dramatic increase of big data generated from complex network systems, such as social networks and large scale sensor networks, to find and locate the errors in big data sets becomes quite challenging with normal computing and network systems. Wang et al. [2] provide a classification for errors on social networks based on error scenarios analysis. This classification includes 6 types of common errors with missing data or erroneous data. This work compares the robustness of four node-level network measures, clustering coefficient, network constraint, and centrality. It performs as a good base for developing error finding and detecting techniques for social networks. Social network is a typical instance of complex networks with graph data sets with it. Hence, the error models and types presented in [2] can be extended for the errors in complex network systems.

D. Classification of Error Types in WSN Data Sets

In this paper, we focus on error detection for numeric big data sets from complex networks.. Considering specific feature of numeric data errors[x], there are several unusual data scenarios demonstrated in Fig. 1. The "flat line faults", a time series of a node in a network system keeps unchanged for unacceptable long time duration. The "out of data bounds faults" indicates impossible data values are observed based on some domain knowledge. The "data lost fault" means there are missing data values in a time series during the data generation or communication. The "spike faults" indicates in a time series data items which are totally out of the prediction and normal changing trend.

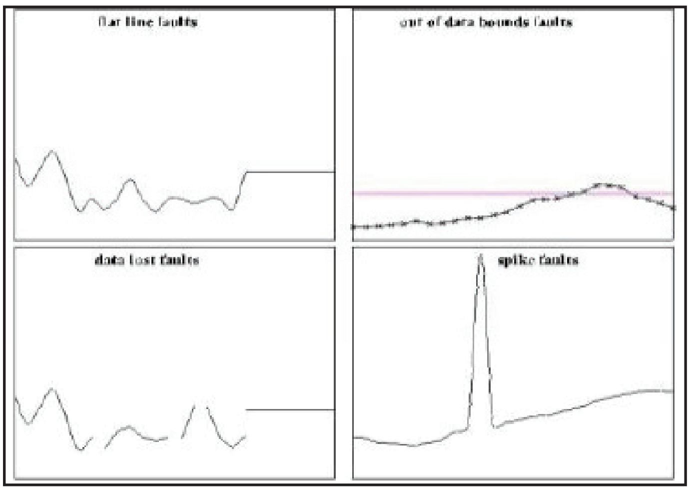

Fig. 1: An Approach for Detecting and Analyzing Errors of Big Sensor Data on Cloud

In the implementation the sensor data is been collected through the simulation, the data sets through the simulation using the simulator. The sensor data is given to the preprocessor to process the data. By using the error tracker I would track the errors. The algorithms used to track all four errors are mentioned below. The first used algorithm is for detecting the missing error algorithm.

IV. Proposed Approach

We intend to develop a novel error acknowledgment approach by abusing the enormous stockpiling, versatility and calculation constrain of cloud to recognize errors in big data sets from sensor frameworks. Snappy acknowledgment of data errors in big data with cloud stays testing particularly, how to use the calculation compel of cloud to quickly find and discover errors of centers in WSN ought to be researched. Cloud Computing is getting the opportunity to be common in light of the fact that it gives an open, versatile, flexible and reconfigurable stage. The proposed error recognizable proof approach in this paper will be established on the request of error sorts. Specifically, nine sorts of numerical data varieties from the standard/errors are recorded and displayed in our cloud error revelation approach. The described error model will trigger the error recognizable proof process. Appeared differently in relation to past error area of sensor framework structures, our strategy on cloud will be made and made by utilizing the colossal data taking care of limit of cloud to enhance error acknowledgment speed and continuous reaction. Our proposed error revelation approach on cloud is especially trimmed for finding errors in big data sets of sensor frameworks. The essential responsibility of our proposed acknowledgment is to fulfill big time execution change in error area without exchanging off error disclosure precision.

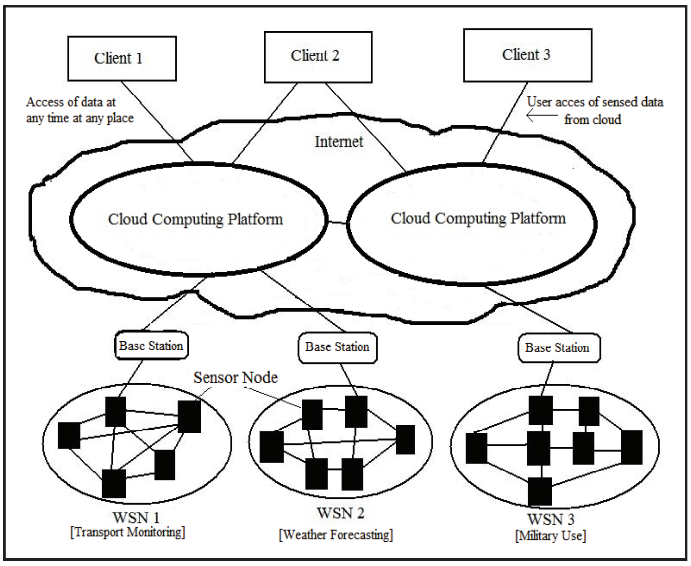

Fig. 2: Proposed System Architecture

Proposed Algorithm

Algorithm -1 Detection of missing error
Step 1 Collect the data sets that the id, time, value.
Step 2 Differentiate data sets for values based on
—#‖. Step 3 Initialize the sensor id=0.
Step 4 for each i=0 to i++.
Step 5 Find the value of diff=time[0]-time[1].
Step 6 for each i=2 to i++.
Step 7 Find the value of dn=time[i]-time[i-1].
Step 8 If dn= diff , no error.

Else missing error.

The detection of the flat line error algorithm is given in algorithm 2 below.

Algorithm-2 Detection of flat line error
Step 1 Initialize a variable time =0.
Step 2 Initialize obstime = -1.

Step 3 for each i=1 to i++.
Step 4 val= value[i]-value[i-1].
Step 5 if val=0.0 time++
obstime = 1
Step 6 Else time=0 Obstime = -1
Step 7 if time is greater than 5 Return flat line error

The detection of the spike error algorithm is given in algorithm 3 below.

Algorithm-3 Detection of spike error
Step 1 Get the value of the sensor.
Step 2 for each i=1 to i++
Step 3 Calculate prev=val[i-1]
Current=val[i] next=val[i+1]
Step 4 if(current>prev&&current>next) Avg=(prev+next)/2
Tim=Current/Avg
Step 5 if(tim>5) Return spike error Step 4 Else no error

The detection of the out of bound error algorithm is given in algorithm 4 below

Algorithm-4 Detection of out of bound error
Step 1 Initialize the value of maxvalue=60.
Step 2 for I=0 to i++
Step 3 if value is greater than max value return out of bound error
Step 4 else return there is no out of bound error.

The entire above mentioned algorithm is implemented to detect the errors from the sensor data and finally corrected.

V. Conclusion and Future Work

Error occurrence in the big sensor data becomes the greatest issue in the real world application which affect the original behavior of the data's that are collected. Detecting errors that are present in the large volume would consume more time complexity. In the existing work it is resolved by comparing the source error patterns with the gathered data entries to find and locate the errors that are present in the environment. The existing system is focused on only detection errors which cannot correct the errors that are detected. This is resolve by using forward errors correction mechanism which can correct the errors that are present in the environment. Also in the proposed system, we propose neural network algorithm to significantly detect the errors and to increases the accuracy. In future, we can develop an advanced algorithm to progress the optimal performance in the given scenario and reduces the error rates greatly. It will improve the higher accuracy in big sensor data applications.

References

[1] Arantxa Duque Barrachina, AislingO'Driscoll,"A big data methodology for categorizing technical support requests using Hadoop and Mahout", Springer Journal of Big Data 2014, 1:1.

[2] Shaojie Tang, Jing Yuan, Cheng Wang, Xiang-Yang Li,,"A Framework for Amazon EC2 Bidding Strategy under SLA Constraints", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 1, January 2014.

[3] AdilFahad, NajlaaAlshatri, ZahirTari, Abdullah Alamri, Ibrahim Khalil, Albert Y. Zomaya, SebtiFoufou, and AbdelazizBouras,"A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis", IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING October 2014.

[4] Chi Yang, Chang Liu, Xuyun Zhang, Surya Nepal, Jinjun Chen,"A Time Efficient Approach for Detecting Errors in Big Sensor Data on Cloud", IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 2, February 2015.

[5] Chang Liu, Jinjun Chen, Laurence T. Yang, Xuyun Zhang, Chi Yang, Rajiv Ranjan, Ramamohanarao Kotagiri, "Authorized Public Auditing of Dynamic Big Data Storage on Cloud with Efficient Verifiable Fine-Grained Updates", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 9, September 2014.

[6] Fan Zhang, Junwei Cao, Kai Hwang, Keqin Li, Samee U. Khan,"Adaptive Workflow Scheduling on Cloud Computing Platforms with Iterative Ordinal Optimization", IEEE Transactions on Cloud Computing 2013.

[7] Shifeng Fang, Li Da Xu, Yunqiang Zhu, JiaerhengAhati, Huan Pei, Jianwu Yan, Zhihui Liu,"An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things", IEEE Transactions on Industrial Informatics, Vol. 10, No. 2, May 2014.

[8] Lin Gu, DezeZeng, Peng Li, Song Guo,"Cost Minimization for Big Data Processing in Geo-Distributed Data Centers", IEEE Transactions on Emerging Topics in Computing October 2014.

[9] Katarina Grolinger, Wilson A Higashino, AbhinavTiwari and Miriam AM Capretz,"Data management in cloud environments: NoSQL and NewSQL data stores", Springer Journal of Cloud Computing: Advances, Systems and Applications 2013, 2:22

[10] Xixian Han, Jianzhong Li, Donghua Yang, Jinbao Wang, Efficient Skyline Computation on Big Data", IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 11, November 2013.

Mis B.Bhavani,B.Tech (IT) in Department of IT,Avanthi'sST. Therassa Institute of engineering &t echnology,Garividi,Vizianagaram.M .Tech(CSE) in Department of CSE Avanthi Institute of Engineering &Technology,Vizianagaram,She Holds a B.Tech ceritificate in Information technology.Attended IT Workshops.One of the active member in student committees.Done Microsoft certification in "Dot NET".Done some social activities, medical camps behalf of trust.Areas of interests are Web technology, Object oriented analysis and design,Data Communication systems, Computer Networks.
He believes in the wording of APJ Abdul Kalam
"Dreams is not what you see in sleep,is the thing which doesn't let you sleep".

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

V. Mahesh, M.Tech (Ph.D), is working as an Assistant Professor in Department of CSE, Avanthi Institute of Engineering & Technology, Vizianagaram, AP, INDIA. He done his M.Tech post graduate in Computer Science & Engineering from JNTU Kakinada. He attended several seminars and workshops. He published several International Papers which shows his zeal towards research.
He believes in the wordings of Swami Vivekananda
"Arise! Awake! and stop not until the goal is reached"

Dr. A.ChandraSekhar, Professor & HOD of Department of Computer Science And Engineering Avanthi Institute of Engineering &Technology, Vizianagaram, AP, Affiliated to JNTU Kakinada. . He attended several seminars and workshops. He published several International Papers which shows his zeal towards research.

He believes in the wordings of Albert Einstein:
"The True Sign Of Intelligence Is Not Knowledge But Imagination"