DETECTING AND LOCATING OF ERRORS OF DATA IN CLOUD

[1] K.SOWMYA REDDY,

[2] V.DIVYA RAJ, ASSISTANT PROFESSOR (M.Tech),

[3] DR. M.SEETHA, PROFESSOR & HOD (Ph.D),

[1]sowmyadandy@gmail.com, [2] divya.chunduri@gmail.com, [3] smaddala2000@yahoo.com

[1, 2, 3] G.NARAYANAMMA INSTITUTE OF TECHNOLOGY &SCIENCE, SHAIKPET, HYDERABAD.

Abstract-- Big sensor data is huge amount of data in both industry and scientific research application in which it generate high quantity of data. Cloud computing provides a special platform to support this challenge as it provides a flexible massive data, storage, and different software services in a scalable manner at low cost. Different technique has been developing in recent years for processing sensor data on cloud, such as sensor-cloud. But, these techniques do not provide efficient support on fast detection and locating of errors in big sensor data. For fast error detection in big sensor data sets, in this system, we develop a novel based approach for both detection and localization.Which gives the full feature of cloud platform and the network feature of Wireless sensor network(WSN). Firstly, it classifies a set of sensor data error types and then defines it. Specifically, in proposed system, the error detection is based on the scale-free network topology and most of detection operations can be conducted in clustered form not a whole big data set. This approach can reduce the time for detection of the error and also location of the error in big data sets and main advantage is fixing the errors.

Keywords-- Big data, data abnormality, time efficiency, sensor networks, complex network systems.

I. INTRODUCTION

The cause of data explosion in the present era the biggest challenge faces is processing of the big data. Since big data is collection of data sets and it so complex to process it as the data keeps on exploring. The traditional approach of human cognitive process which includes datasets which is beyond the ability to process the data in tolerable elapsed time which would be a major drawback since datasets keeps on accumulating day by day and becomes difficult task to process it. One of the major and important characteristic of big data is volume, velocity, value, veracity and variety. The big data sets can from any base such as meteorology, complex physics simulations, biological study and environmental research.

One important source of data set is collected by wireless sensor network (WSN). The WSN have feature of enhancing the ability of monitoring and interact with physical environment. Since there is corruption and lose of data due to presence of WSN in hardware inaccuracies in the node. It is necessary for data to be received clean and accurate. There is a need of effective detection and also cleaning of sensor big data is a major challenging and requires innovative solutions. WSN with the cloud can be called as complex network systems. As the complex network increases the data in accuracy and error has become an issue in real network application.

WSN big data error detection usually requires real time processing and also storage for massive sensor data which would also use the complex error model to detect the event of abnormality. In this paper we aim to develop a approach by having massive storage, scalability and also having computation power to detect error in big data sets from sensor data. The proposed error detection approach in this paper is by detecting the types of errors.

II. RELATED WORK

As there was increase in the data, the processing of

the data sets also increased in era of data explosion. The different techniques for processing the data are as follows:

A)Big Data Processing on Cloud:-The cloud computing has a significant feature of big data storage and also interpreting it massive computation power. A design called "stream-as-you-go" is used for the processing the data has it is increased through the stream based management data architecture. Map Reduce is also used to process to analyze the incremental data and also process by dividing the data. But all this focus on the workload distribution, scalability and filtering of the data in speed but these approach is not enough for error detection and correction.

B)On- Cloud Processing for WSN:-The WSN are used in different fields such as in environment monitoring, military and also in scientific data collection. Since the variety of data has been collected from different fields through the WSN, there is a need of the sensor cloud for processing of the data. But due the nature of big data with the feature of volume, variety and velocity it is difficult to process the data in the complex network system. By effect of this using the cloud sensor it is difficult to develop the time efficient approach to detect the errors in the data sets and also to debug the complex network system in real time.

C)Data Error Detection in Sensor Network and Complex Network:- The data error is unavoidable in real world complex network system. As there was dramatic increase in big data generated to locate the error was a quite challenging task with normal computing and network system. Wang et al has proposed a classification of errors based on scenario analysis. It performed well in finding the errors detecting the errors. This work compares the robustness of four node network only, clustering co-efficient and centrality and hence can be extended for the complex network system.

Mukhopadhyay proposed error correction method for the WSN. It was performed on the intelligent sensor network. The intelligent sensor network correction of the error was faster. But as the big data sets increased its processing capability and time performance was extremely limited when the data was increasing.

In our approach we would detect the errors in the complex network by improve the time efficiency and faster detection and location of the errors.

III. PROPOSED WORK

In this paper we discuss about reducing the time for analyzing the errors and fixing of the error in the big data sets. Before applying this approach we need to classify the types of errors. The errors can be classified based on the text file data. Since the errors are so common in complex network system. The architecture of the proposed system is follows which clearly specify the structure and flow of the system.

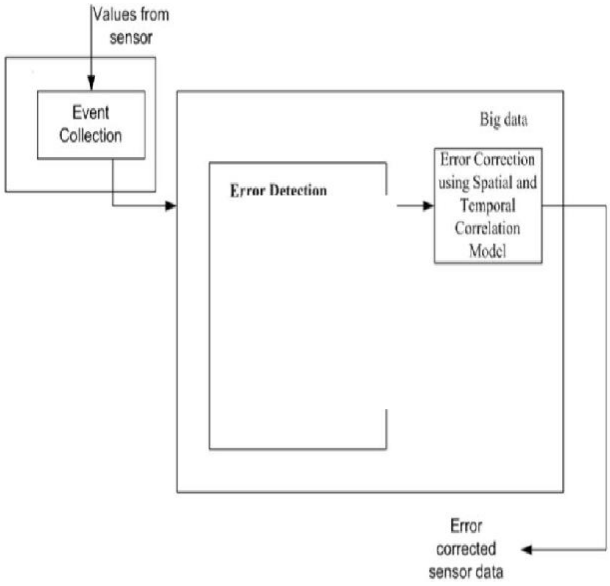

Fig-1 Architecture diagram

The errors are classified as follows

1)Types of Errors

A) The flat line faults shown in fig-2 indicate that nodes in the network kept unchanged for unacceptable time series and for long duration. In real world application the transmitted data will have small

changes over time flow.

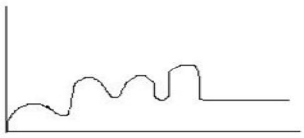

Fig-2 Flat line fault

B) The out of bound faults shown in fig-3, the value of the data can be observed based on the domain knowledge that is gained in general. In real world applications if a threshold is fixed for a wave, if any wave beyond that it is treated as out of bound error.

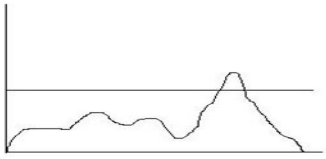

Fig-3 Out of bound fault

C) The data lost fault shown in fig-4, it means that there is missing of data over time during the data is generated and exchanged process which requires data cleaning.

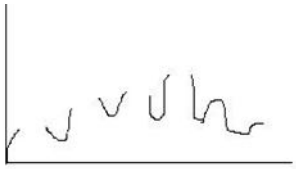

Fig-4 Data lost fault

D) The spike fault is shown in fig- 5 the spike error indicates that in a time series data items which would be out of prediction suddenly and normal over the time series

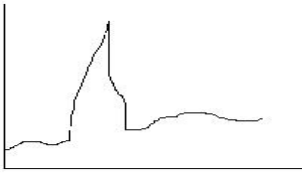

Fig-5 Spike fault

2) ERROR DETECTION

There are different types of work done on the big data set using different types of classification algorithm. In big data set a complex network is represented on which error detection algorithm is applied. The error detection algorithm firstly analysis whole big data set then it filters out it. Clustering process is done on given big data set. Clustering can significantly reduce the time and cost for detecting errors and finds the exact location of error on given data set. Efficiently locate the error on big data set and then after recover the corrupted files. This is the overall process using these errors detection strategies are done and find the errors and locate it using different time efficient technique. This system aims to develop a novel error detection approach for the big data storage, scalability and computation power of cloud to detect and locate the errors in big data sets from sensor networks.

In proposed approach, error detection is conducted in limited time period on big data set. Clustering is done over the big data. It does not consider the whole data set instead it uses clustering. Therefore the error detection and location process can be done fast. Furthermore the system uses the User define cloud known as U-cloud. Proposed system can easily minimize the time for error detection and location in big data sets generated by large scale sensor network systems with properly defined error detection accuracy.

3) ERROR LOCALIZATION

The importance of localization information arises from several factors, many of which are related only to WSNs. These factors include the identification and correlation of gathered data, node addressing, management and query of nodes localized in a determined region, evaluation of nodes' density and coverage, energy map generation, geographic routing, object tracking, and other geographic algorithms. All of these factors make localization systems a key technology for the development and operation of WSNs.

Finding location of error is important after detection. Error localization algorithm is used to locate position and source of error in network . Error localization algorithms helps in diagnosing the root cause of error. Data recovery plays important role in both error detection and localization in complex network systems.

IV. IMPLEMENTATION

In the implementation the data is been collected through the simulation, the data sets through the simulation using the simulator. The data is given to the pre processor to process the data. By using the error tracker it 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.

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 data and finally corrected.

V. 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 detection and location 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 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.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, ―A Survey on Sensor Network‖, IEEE Comm. Mag., vol. 40, no. 8, 2002.

[2] Wireless Sensor Networks, Technology, Protocols and Application by K. Sohraby, D. Minoli and T. Znati

[3] S. Mukhopadhyay, D. Panigrahi, and S. Dey Data Aware, Low Cost Error Correction for Wireless Sensor Network,‖ proc. IEEE Wireless Comm. And Networking Conf.Sensor Cloud,‖ http://www.sensorcloud.com

[4] S. Sakr, A.Liu, D. Batista, and M. Alomari. A Survey of Large Scale Data Management Approaches in Cloud Environment,‖ IEEE Comm. Survey.

[5] M.C. Vuran and I. Akyildiz, ―Spatial Correlation based on Collaborative Medium Access Control in Wireless Sensor Network.

[6] W. Dou, X. Zhang, J. Liu, and J. Chen. HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big DataApplications.IEEE Trans. Parallel and Distributed Systems, 2013

[7] Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, ―A View of Cloud Computing,‖ Comm. the ACM, vol. 53, no. 4,pp. 50-58, 2010.

[8] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud Computing and Emerging it Platforms: Vision, Hype, and Reality for Delivering Computing As the 5th Utility,‖ Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009

[9] L. Wang, J. Zhan, W. Shi, and Y. Liang, ―In Cloud, Can Scientific Communities Benefit from the Economies of Scale?‖ IEEE Trans. Parallel and Distributed Systems, vol. 23,, no. 2, pp. 296-303, Feb