AN EFFICIENT APPROACH FOR DETECTING ERRORS IN BIG SENSOR NETWORK DATA ON CLOUD

S. Md Ismail, M. Tech, Associate Professor, Department of CSE, AI – Habeeb College of Engineering & Technology, Chevella, Hyderabad.

Mr. Mohd Anwar Ali, M. Tech, Associate Professor & HOD, Department of CSE, AI – Habeeb College of Engineering & Technology, Chevella, Hyderabad.

Syeda Hafsa Hashmi, Post Graduate Student, Department of CSE, AI – Habeeb College of Engineering & Technology, Chevella, Hyderabad.

ABSTRACT – The data is being generated from all types of IT and Non-IT devices and from many environments. This data from all the directions can be said a Big Data. Big data is playing a key role in scientific and research applications and it is being generated with a high amount of volume and high velocity. This kind of data is difficult to process by using traditional on-hand data processing tools and applications. At Present, A Cloud Computing platform offers a flexible stack of computing, huge storage, and many services in scalable manner with low pricing schemes. There are some techniques to process sensor data on cloud for the fast detection and localization of errors in the large size of sensor datasets. It is good that we have techniques to process a big sensor data on cloud, but these techniques do not provide an efficient support on fast detection and localization of the errors in large size sensor datasets. In this paper, we introduce a new data error which exploits the maximum computing potential of cloud platform and a network feature of WSN. Initially we classify and define error type and then according to the classification a network feature of clustered wireless sensor network will be introduced and analyzed to support a fast error detection and localization.

I. INTRODUCTION

Recent years have seen several tremendous advances in the hardware technology such as the development of miniaturized sensor devices, GPS-enabled devices, pedometers, and accelerometers, which may be accustomed to collect various kinds of information. This has cause a deluge of tremendous amounts of real-time data, which may be well-mined for a range of analytical insights. The cost of sensing hardware has been systematically taking place over the past few years. Moreover, several data collecting technologies such as RFID are enabled during a very cost-efficient means, as a result of which the scale of the gathering method has become huge.

The sensor data is being produced in the context of a wide variety of applications such as the following:

- From a wide variety of GPS enabled mobile devices.
- Because of low cost of RFID tags, the tremendous volume of RFID data is being stored.
- Many of the military applications make use of various types of sensors to track the illegal events or activities.
- Sensors are also deployed in the context of many environmental applications like detecting weather and the climate situations pollution tracking and so on.

Big Data in general is a collection of the large and complex data sets that it becomes difficult to process with on hand database management systems or with the traditional data processing applications. Cloud computing provides a promising platform for large processing with powerful computation capability, storage, measurability, resource utilize and low price, and has attracted important
attention in alignment with huge data. One of necessary source for scientific huge data is that the data sets collected by wireless sensing device networks (WSN). Wireless sensor networks have potential of considerably enhancing people’s ability to observe and move with their physical environment. Big data set from sensors is commonly subject to corruption and losses owing to wireless medium of communication and presence of hardware inaccuracies within the nodes. For a WSN application to deduce an applicable result, it is necessary that the information received is clean, accurate, and lossless. However, effective detection and cleanup of sensing element huge data errors may be a difficult issue demanding the innovative solutions.

Some work has been executed for enormous data analysis and error detection in elaborate networks together with intelligence sensors networks. There are additionally some works involving problematic network techniques knowledge error detection and debugging with on-line data processing approaches. On the grounds that these techniques were not designed and developed to deal with massive data on cloud, they have been unable to cope with current dramatic increase of knowledge size. For example, when the big data sets are encountered, the previous offline methods for the error detection and debugging on a single computer may take a long time and lose real time feedback. Because those offline methods are generally based on learning or mining, they regularly introduce high time cost during the processing of dataset training and pattern matching.

Cloud computing infrastructure is turning into widespread as a result of it provides an open, flexible, ascendible and reconfigurable platform. The projected error detection approach during this paper are going to be supported the classification of error types. Specifically, nine kinds of numerical information abnormalities/errors are listed and introduced in our cloud error detection approach. The outlined error model can trigger the error detection method. Compared to previous error detection of sensor network systems, our approach on cloud are going to be designed and developed by utilizing the large processing capability of cloud to boost error detection speed and real time reaction. Additionally, the design feature of complicated networks also will be analyzed to mix with the cloud computing with a additional economical method. Based on current analysis literature review, we tend to divide complicated network systems into scale-free sort and non scale-free kind. Sensor network could be a reasonably scale-free complicated network system that matches cloud quantifiability feature. Our planned error detection approach on cloud is specifically cut for finding errors in massive data sets of sensor networks. The main contribution of our planned detection is to attain significant time performance improvement in error detection while not compromising error detection accuracy.

II. RELATED WORK

To deal with various challenges of tremendous data, research works will also be observed intensively from the data base view. However, the quandary can also be additionally mentioned from the viewpoint of parallel programs and cloud. On this section, associated literature for tremendous knowledge processing on cloud, and knowledge error detection for complex network systems will probably be reviewed and compared.

At present, some work has been carried out for processing significant information with cloud. Amazon EC2 infrastructure as a carrier is a common cloud headquartered distributed process for giant knowledge processing. Amazon S3 supports disbursed storage. Map Reduce is adopted as a programming mannequin for large information processing over cloud computing. Plenty of recent research has investigated the problems of processing incremental information on cloud. Kienzler et al designed a “circulate-as-you-go” process to entry and system on incremental knowledge for data-intensive cloud purposes by way of a flow-established data
administration structure. The extension of the traditional Hadoop framework to improve a novel framework named Incoop via incorporating a number of tactics like task partition and memorization-conscious schedule. Olston et al. gift a steady workflow approach referred to as Nova on top of Hadoop by means of stateful incremental knowledge processing. Map Reduce has been generally revised from a batch processing framework right into an extra incremental one to investigate enormous-volume of incremental knowledge on cloud. It is a framework for processing parallelizable problems across giant data sets making use of a tremendous quantity of computer systems (nodes), at the same time referred to as a cluster in which all computers (nodes) are on the identical nearby community and use an identical hardware; or a grid in which the nodes are shared across geographically and administratively disbursed programs. It could actually style a peta byte of data in only a few hours. The parallelism additionally presents some possibility of recuperating from partial failure of servers or storage during the operation.

Sensor-Cloud is a specified sensor information storage, visualization and faraway administration platform that leverages robust cloud computing applied sciences to provide high-quality information scalability, speedy visualization, and person programmable evaluation. Initially, sensor-cloud was once designed to help lengthy-time period deployments of Micro Strain wi-fi sensors. However nowadays, sensor-cloud has been developed to help any web-linked 1/3 celebration gadget, sensor, or sensor network via a simple Open Data API. Sensor-Cloud will also be useful for a form of functions, principally the place information from colossal sensor networks desires to be collected, seen, and monitored remotely.

It may be concluded that current information error detection techniques for complex network programs focus on in-network detecting with intelligent nodes or offline evaluation on the root. They ignore the scalability, massive resource and robust computation ability supplied via cloud. The proposed approach in this paper aims to address this difficulty by means of utilizing the inherent features of cloud computing to have an understanding of rapid error detection on cloud. Compared to the previous sensor data error detection and localization procedure, complicated community topology facets will probably be explored with the computation energy of cloud for error detection affectivity, scalability and low price.

III. FRAMEWORK

A. Error and Abnormality Classification

Under the theme of the large data sets from real world complicated networks, there are primarily two sorts of knowledge generated and exchanged among networks. (1) The numeric information sampled and exchanged between network nodes like sensing element network sampled data sets. (2) The text files and information logs generated by nodes like social network data sets. In this paper, our analysis can specialize in the error detection for numeric huge data sets from advanced networks.

Fig.1. Error Scenarios from sensor network systems data

Considering a particular feature of numeric data errors, there are various abnormal data scenarios given in figure.

Flat line fault: These faults indicate a time series of a node in a network system keeps unchanged for unacceptable long time period. In real world applications, sampled data and transmitted data is always have slight changes with a time flow.
**Out of data bounds faults:** indicates impossible data values are observed by some domain knowledge. In context real world applications, if a temperature of water is reported as 300 degree Celsius, it can be treated as a data fault directly.

**Data lost faults:** means if there are missing data values in a time series in a data generation or in a communication. The time series with “data lost fault” generally needs data cleaning.

**Spike fault:** These indicate in a time series data items which are totally out of prediction and normal changing trend.

Because the given four types of errors can take place both at data generation and data exchange stages, the types of errors can also be categorized into node side and edge side separately. In this paper, we propose a two-phase approach to conduct a computation required in entire process of error detection and localization.

**B. Error Detection Algorithm**

In the phase of error detection, mainly there are three inputs for the error detection algorithm. The first one is a graph of network. The second one is a total collected data set D and the third one is the defined error patterns p. The output result of the error detection algorithm is the error set D.

**C. Error Localization Algorithm**

After the process of error pattern matching and error detection, it is necessary to locate the position and source of a detected error in an original Wireless Sensor Network graph G(V, E). Input of the Algorithm 2 is an original graph of a scale-free network G (V, E), and an error data D from the Algorithm 1. The output of algorithm 2 is G’(V’, E’) which is a subset of the G to indicate error location and source.

**IV. FURTHER ENHANCEMENT**

In the present we have worked toward the error classification. Obviously, in the context of this paper we have studied about various new error types and classified them as node side and edge side errors. In this paper, we have mainly worked on fast error detection and placing of errors. In future the work is motivated to correct errors on the basis of detection and localization, a big data cleaning, a data recovery and false positive integration of the large amount of data of efficient and fast processing.

**V. EXPERIMENTAL RESULTS**

To verify an efficiency of a time and the effectiveness of our proposed approach for detecting the errors in big data with the cloud, experiments are conducted on real world sensor dataset. There are three major purposes for this experiment. 1) Describes that a significant time-saving is achieved in terms of detecting errors from the complex network big data sets. 2) Demonstrate an effectiveness of our proposed error detection approach in terms of various error types. 3) It demonstrates that the false positive ratio of the proposed error detection algorithm is limited in a small value. The screens given below show the experimental results of our work.
VI. CONCLUSION

In order to notice errors in large data sets from sensor network systems, a unique approach is developed with cloud computing. First of all error classification for a large data sets is presented. Secondly, the correlation between sensor network systems and also the scale-free complicated networks are introduced. In line with every error type and also the features from scale-free networks, we have projected a time-efficient strategy for detection and locating errors in huge data sets on cloud. With the experiment results from our cloud computing surroundings, it's incontestable that 1) the proposed scale-free error detection approach will considerably reduce the time for quick error detection in numeric big data sets, and 2) proposed approach achieves similar error selection ratio to non-scale-free error detection approaches.

References


