Optimized Error Detection in Cloud User for Networking Services

[1] T. A. Sharanya
M.Sc., M.Phil.
Department of Computer Science,
Selvamm Arts and Science College
(Autonomous)
Namakkal (Tk) (Dt) – 637003.

[2]Mrs. D. Ananthanayaki
M.C.A., M.Phil.
Assistant Professor, Department of Computer Science,
Selvamm Arts and Science College
(Autonomous)
Namakkal (Tk) (Dt) – 637003.

[3]Mrs. K. K. Kavitha
M.C.A., M.Phil., SET., (Ph.D).,
Vice Principal, Head of the Department of Computer Science,
Selvamm Arts and Science College
(Autonomous)
Namakkal (Tk) (Dt) – 637003.

Abstract: Big sensor data is prevalent in both industry and scientific research applications where the data is generated with high volume and velocity it is difficult to process using on-hand database management tools or traditional data processing applications. Cloud computing provides a promising platform to support the addressing of this challenge as it provides a flexible stack of massive computing, storage, and software services in a scalable manner at low cost. Some techniques have been developed in recent years for processing sensor data on cloud, such as sensor-cloud. However, these techniques do not provide efficient support on fast detection and locating of errors in big sensor data sets. For fast data error detection in big sensor data sets, in this paper, we develop a novel data error detection approach which exploits the full computation potential of cloud platform and the network feature of WSN.

*****

I. INTRODUCTION TO CLOUD

Recently, we enter a new era of data explosion which brings about new challenges for big data processing. In general, big data is a collection of data sets so large and complex that it becomes difficult to process with on hand database management systems or traditional data processing applications. It represents the progress of the human cognitive processes, usually includes data sets with sizes beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time. Big data has typical characteristics of five 'V's, volume, variety,velocity, veracity and value. Big data sets come from many areas, including meteorology,connectomics, complex physics simulations, genomics, biological study, gene analysis and environmental research. According to literature, since 1980s, generated data doubles its size in every 40 months all over the world. In the year of 2012, there were 2.5 quintillion (2.5 _ 1018) bytes of data being generated every day. Hence, how to process big data has become a fundamental and critical challenge for modern society. Cloud computing provides a promising platform for big data processing with powerful computation capability, storage, scalability, resource reuse and low cost, and has attracted significant attention in alignment with big data.

One of important source for scientific big data is the data sets collected by wireless sensor networks (WSN). Wireless sensor networks havepotential of significantly enhancing people's ability to monitor and interact with their physical environment. Big data set from sensors is often subject to corruption and losses due to wireless medium of communication and presence of hardware inaccuracies in the nodes. For a WSN application to deduce an appropriate result, it is necessary that the data received is clean, accurate, and lossless.

II. BIG DATA IN CLOUD COMPUTING ENVIRONMENT

The successful paradigm for the service orientedprogramming is the cloud computing. It has revolutionizedthe way of computing infrastructure's abstraction andusage. The elasticity, pay per use, low upfront investment,transfer of risks are few of the major enablingcharacteristics that makes the cloud computing theubiquitous platform for deploying economically feasibleenterprise infrastructure settings. Distributed databases hadbeen the boon of vision for research for few decades. Butchanges in the data patterns and applications has made wayfor the new type of storage called key value storage whichare now being widely used by various enterprises. In thedomain of Map reduce and open source implementationof the same known as the Hadoop has been used bymajority of the industry and academics. Hadoop increasesthe usability and performance.HDFS has become a very helping tool to maintain and store the complex data.Big data has becoming more available and understandableto computers. What is big data? The question arrives. Bigdata is the representation of progress of the humancognitive processes, usually includes data sets with sizesthat is beyond the current technology's capability.

BIG DATA

Big data is a word used for description of massiveamounts of data which are either structured, semi structuredor unstructured. The data if it is not able to be handled bythe traditional databases and software technologies then wecategorize such data as big data. The term big data isoriginated from the web companies who used to handleloosely structured or unstructured data. The big data isdefined using three v's.1) Volume: many factors contribute for the increase involume like storage of data, live streaming etc.2) Variety: various types of data is to be supported.3) Velocity: the speed at which the files are created andprocesses are carried out refers to the velocity.

Big Data classify them with respect to their strategies including bigdata management platform, distributed file system, big datastorage, MapReduce application and optimization. However,maintaining and processing these large-scale data sets istypically beyond the reach of small businesses and it isincreasingly posing challenges even for large companies andinstitutes. Finally, we discuss the open issues and challengesin processing big data in three important aspects: big datastorage, analysis and security.

OPEN SOURCE CLOUD PLATFORM

The main idea behind data center is to leverage thevirtualization technology to maximize the utilization ofcomputing resources. Therefore, it provides the basic ingredientssuch as storage, CPUs, and network bandwidthas a commodity by specialized service providers at low unitcost. For reaching the goals of big data management, most ofthe research institutions and enterprises bring virtualizationinto cloud architectures. Amazon Web Services (AWS),Eucalptus, Opennebula, Cloudstack and Openstack are themost popular cloud management platforms for infrastructureas a service (IaaS). AWS9 is not free but it has hugeusage in elastic platform. It is very easy to use and onlypay-as-you-go. The Eucalyptusworks in IaaS as anopen source. It uses virtual machine in controlling andmanaging resources. Since Eucalyptus is the earliest cloudmanagement platform for IaaS, it signs API compatibleagreement with AWS. It has a leading position in theprivate cloud market for the AWS ecological environment.OpenNebula has integration with various environments.It can offer the richest features, flexible ways and betterinteroperability to build private, public or hybrid clouds.OpenNebula is not a Service Oriented Architecture (SOA)design and has weak decoupling for computing, storage andnetwork independent components.

CloudStack is an opensource cloud operating system which delivers public cloudcomputing similar to Amazon EC2 but using users' ownhardware. CloudStack users can take full advantage of cloudcomputing to deliver higher efficiency, limitless scale andfaster deployment of new services and systems to the enduser.At present, Cloud Stack is one of the Apache opensource projects. It already has mature functions. However,it needs to further strengthen the loosely coupling and componentdesign. OpenStack11 is a collection of open sourcesoftware projects aiming to build an open-source communitywith researchers, developers and enterprises. People in thiscommunity share a common goal to create a cloud that issimple to deploy, massively scalable and full of rich features.The architecture and components of Open Stack are straightforwardand stable, so it is a good choice to provide specificapplications for enterprises. In current situation, Open Stackhas good community and ecological environment. However,it still has some shortcomings like incomplete functionsand lack of commercial supports.

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. 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 can be extended for the errors in complex network systems.

OVERVIEW OF THESIS

A new era of data explosion which brings about new challenges for big data processing. In general, big data is a collection of data sets so large and complex that it becomes difficult to process with on hand database management systems or traditional data processing applications. It represents the progress of the human cognitive processes, usually includes data sets with sizes beyond the ability of current technology, method andtheory to capture, manage, and process the data within a tolerable elapsed time. Big data has typical characteristics of five 'V's, volume, variety, velocity, veracity and value. Big data sets come from many areas, including meteorology, connectomics, complex physics simulations, genomics, biological study, gene analysis and environmental research.

According to generated data doubles its size in every 40 months all over the world. Hence, how to process big data has become a fundamental and critical challenge for modern society. Cloud computing provides a promising platform for big data processing with powerful computation capability, storage, scalability, resource reuse and low cost, and has attracted significant attention in alignment with big data.One of important source for scientific big data is the data sets collected by wireless sensor networks (WSN). Wireless sensor networks have potential of significantly enhancing people‟s ability to monitor and interact with their physical environment. Big data set from sensors is often subject to corruption and losses due to wireless medium of communication and presence of hardware inaccuracies in the nodes. For a WSN application to deduce an appropriate result, it is necessary that the data received is clean, accurate, and lossless. However, effective detection and cleaning of sensor big data errors is a challenging issue demanding innovative solutions. WSN with cloud can be categorized as a kind of complex network systems. In these complex network systems, such as WSN and social network, data abnormality and error become an annoying issue for the real network applications. Therefore, the question of how to find data errors in complex network systems for improving and debugging the network has attracted the interests of researchers. Some work has been done for big data analysis and error detection in complex networks including intelligence sensors networks. There are also some works related to complex network systems data error detection and debugging with online data processing techniques.

REVIEW OF LITERATURE

Now-A-Day's Maintaining Or Processing Big Data Is A Critical Challenging Problem. One Of Important Source For Scientific Big Data Is The Datasets Gathered Through The Wireless Sensor Networks (Wsn). For A Wsn Application To Realize An Appropriate Result, It Is Required That The Data Received Is Clean, Accurate, And Lossless. However, Effective Error Detection As Well As Cleaning Of Sensor Big Data Errors Is A Challenging Problem Demanding Novel Solutions. In This Paper, We Propose A Novel Error Detection Approach To Detecting Errors In Big Data Sets From Sensor Networks. This Approach Is A Best Solution To Detecting Errors Even Faster In Big Sensor Data. This Proposed Approach Works Based On Scale-Free Network Topology And Also This Approach Will Be Based On The Classification Of Error Types. By Implementing This Approach We CanAchieve Significant Time Performance Enhancement In Error Detection Without Compromising Error Detection Accuracy[1].

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, sensor-cloud 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 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 long-term 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[2].

One essential quality of cloud computing is in aggregation of resources and data into data centers on the Internet. The present cloud 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 analyze such data with the powerful computational capacity of clouds.

In this way, cloud is now in the phase of expanding from application aggregation and sharing to data aggregation and utilization. To make full use of data, tens of terabytes (TBs)or tens of petabytes (PBs) of data need to be handled and a new type of technology different from ordinary information and communications technology (ICT) is required.

III. BACK GROUND OF THE STUDY

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.

We aim to develop a novel error detection approach by exploiting the massive storage, scalability and computation power of cloud to detect errors in big data sets from sensor networks.Fast detection of data errors in big data with cloud remains challenging. Especially, how to use the computation power of cloud to quickly find and locate errors of nodes in WSN needs to be explored.

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. 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 can be extended for the errors in complex network systems. Xiong proposed an approach which can be used to detect the text dataerrors in data sets of social network. 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. Because the work is in- network fast error detection by intelligent sensors, its processing capability and time performance are extremely limited when encountering big data sets. Similar work can also be conducted with the consideration of data awareness and low cost according to the description.Analyses the location errors in sensor networks. The primary goal of this location error analysis is to demonstrate the practical use of the location errors for optimal resource consumption. A detailed study of sensor faults that occur in deployed sensor net- works and a systematic approach to model these faults.

STATEMENT OF THE PROBLEM

To address various challenges of big data, research works can be found intensively from the database view. However, the problem can be also discussed from the perspective of parallel systems and cloud. In this section, related literature for big data processing on cloud, and data error detection for complex network systems will be reviewed and compared.WSN big data error detection commonly requires powerful real-time processing and storing of the massive sensor data as well as analysis in the context of using inherently complex error models to identify and locate events of abnormalities. We aim to develop a novel error detection approach by exploiting the massive storage, scalability and computation power of cloud to detect errors in big data sets from sensor networks. Some work has been done about processing sensor data on cloud.

RESEARCH METHODOLOGY

SYSTEM ANALYSIS

MapReduce is adopted as a programming model for big data processing over cloud computing. Plenty of recent research has investigated the issues of processing incremental data on cloud. A “stream-as-you-go” approach to access and process on incremental data for data-intensive cloud applications via a stream-based data management architecture. The extension of the traditional Hadoop framework to develop a novel framework named Incoop by incorporating several techniques like task partition and memorization-aware schedule. A continuous workflow system called Nova on top of Pig/Hadoop through stateful incremental data processing. MapReduce has been widely revised from a batch processing framework into a more incremental one to analyze huge-volume 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 and use similar hardware; or a grid in which the nodes are shared across geographically and administratively distributed systems. It can sort a petabyte of data in only a few hours. The parallelism also provides some possibility of recovering from partial failure of servers or storage during theoperation. According to the above literature, most of current big data processing and analysis techniques on cloud focus on the workload distribution, scalability, data filtering speed, and query accuracy. However, there is not enough work dedicated to the issue of error detection and correction for big data sets with cloud computing.

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, sensor-cloud 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 network 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 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 long-term 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 OpenData API.

SYSTEM IMPLEMENTATION

BACKGROUND WORK

Many systems in nature can be described as large networks (nodes or vertices connected by links or edges): Friendship networks, Social networks, computer networks, Internet, metabolic networks, power grids, scientific citations, neural networks and large scale sensor networks. Network analysis has been troubled by the issue of measurement of error for a long time. Before deploying an error detection approach on cloud, the error models for big data sets from wireless sensor network systems perspective should be presented first.

Under the theme of the big data sets from real world complex networks, there are mainly two types of data generated and exchanged within networks. (1) The numeric data sampled and exchanged between network nodes such as sensor network sampled data sets. (2) The text files and data logs generated by nodes such as social network data sets. In this paper, our research will focus on the error detection for numeric big data sets from complex networks. In the previous work, the errors of complex networks can be classified as six main types for both numeric and text data.

IV. TECHNIQUES AND ALGORITHM

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 the subset of the G to indicate the error location and source.

The BCH decoding is complicated because it has to locate and correct the errors. Suppose we have a received codeword r(x)=r0+r1x+r2x2+….+rn-1xn-1, then r(x)=v(x)+e(x), where, v(x) is correct codeword and e(x) is the error. First, we must compute a syndrome vector s=(s1,s2,…,s2t), which can be achieved by calculating r.HT, where, H is parity-check matrix and can be defined as:

$$H = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ 1 & \alpha^2 & (\alpha^2)^2 & \dots & (\alpha^2)^{n-1} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 1 & \alpha^{2t} & (\alpha^{2t})^2 & \dots & (\alpha^{2t})^{n-1} \end{bmatrix}$$

Here, α is the element of the GF field and can be located in the GF table. The location numbers for the errors will be achieved by finding roots of σ(x).

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.

Algorithm

Encode

For i=1…n do

For j=1 …k do

T_j [h_j ((x_i, i))]. Keysum = (x_i, i)

T_j [h_j((x_i, i))].valueSum = Check((xi; i))

Decode

for i = 1….. n do

for j = 1 …. k do

T_j [h_j((y_i, i))]:keySum ^= (y_i, i).

T_j [h_j((y_i, i))]:valueSum ^= Check((y_i, i)).

While ∃ a, j with (Tj [a]:keySum ≠ 0) and

(Tj [a]:valueSum == Check(Tj [a]:keySum)) do

(z, i) = Tj [a].keySum

if z ≠yi then

set y_i to z when decoding terminates

for j = 1 …. k do

T_j [hj((z; i))].keySum ^= (z, i).

Our approach works by automatically identifying subsets of an application's data that are not directly used in computation, and exposing them to the cloud only in encrypted form. We present a technique to partition encrypted data into parts that are accessed by different sets of users (groups). Intelligent key assignment limits the damage possible from a given key compromise, and strikes a good tradeoff between robustness and key management complexity. We present a technique that enables clients to store and use their keys safely while preventing cloud-based service from stealing the keys. Our solution works today on unmodified web browsers.

V. IMPLEMENTATION

A motivating scenario is illustrated in where an online health service provider, e.g., Microsoft HealthVaule , has moved data storage into cloud for economical benefits. Original datasets are encrypted for confidentiality. Data users like governments or research centres access or process part of original datasets after anonymization. Intermediate datasets generated during data access or process are retained for data reuse and cost saving in cloud database. Two independently generated intermediate datasets in Fig.1 are anonymized to satisfy 2-diversity, i.e., at least two individuals own the same quasi-identifier and each quasi-identifier corresponds to at least two sensitive values. Knowing that a lady aged 25 living in 21400 (corresponding quasi-identifier is ⟨214 ∗, female, young⟩) is in both datasets, an adversary can infer that this individual suffers from HIV with high confidence if (a) and (b) are collected together.

VI. RESULT AND DISCUSSION

These five types of data errors are generated equally. Hence, the percentage of each type of errors is 20 percent from the total imposed errors for testing. The first imposed error type is the flat line error. The second imposed error type is out of bound error. The third imposed error type is the spike error. The forth imposed error type is the data lost error. Finally, the aggregate & fusion error type is imposed. By imposing the above listed five types of data error types, the experiment is designed to measure the error selection efficiency and accuracy during the on-cloud processing of data set.

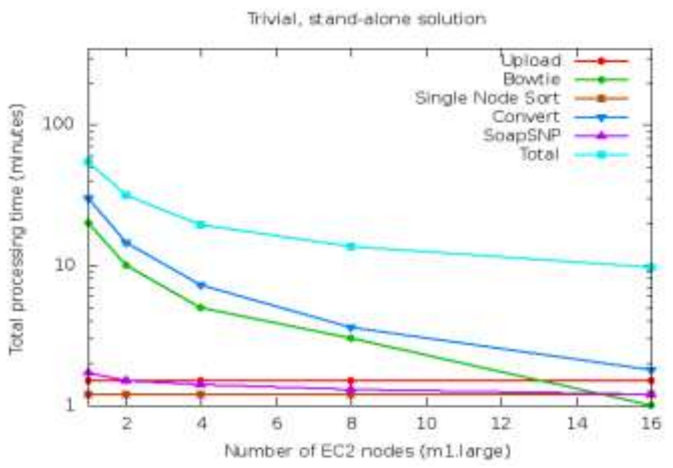
PERFORMANCE ANALYSIS

Method	Advantages	Limitation
Two Phase Top Down Specialization	Parallelization & solves scalabilty problem.	Scalability problem
MRTDS Framework	Reduce Communication Traffic	Data splitting cause transmission overhead.

The investigation of scalability problem of large-scale data anonymization by TDS, and presents a scalable two-phase TDS approach using MapReduce on cloud. The data set are divided and anonymized in parallel in the first phase, producing intermediate results. Map Reduce applied on cloud for data anonymization and deliberately designed a group of innovative MapReduce jobs to concretely accomplish the specialization computation in a highly scalable way. With this approach the scalability and efficiency of TDS are improved significantly over existing approaches.

The Trivial Solution: This solution is implemented based onstandard UNIX commands rsync, scp, split, cat, sort as well ascommand line tools bowtie and soapsnp. Assuming the inputread file is located on a local system, rsync is used to copythe data to a single cloud node. rsync provides online datacompression, which helps us reduce the data transfer time.After the data transfer, the read file is split into n partitionsusing the UNIX command split, where n indicates cluster size.After the split, data partitions are transferred to the n clusternodes using scp. In this case, rsync is not used, since withinthe cluster, compression is not necessary due to sufficientbandwidth (more than 200 MBit/s). Using the partitioned readdata, bowtie is invoked in parallel on each cluster node. Bowtieresults are then transferred to a single cloud node to be mergedusing cat and sorted using sort. After the sort, partitions are rebuiltusing split which are then transferred to the cluster nodes.Since Bowtie output and SOAPsnp input are not completelycompatible, we have used a custom-written Java program inorder to convert between these two data formats in a parallelfashion. Subsequently, soapsnp has been invoked in parallelas well. After transferring the results of SOAPsnp to a singlecloud node, we again use the cat command to merge all resultsto a single file, which is finally transferred back to the localclient machine using rsync.

ISSN: 2454-4248
142 – 148



The total processing time for different stagesof our stream-as-you-go implementation as the number of clusternodes is increased. However, the graphs of this experimentshould be interpreted differently than the previous graphs,since, due to the incremental processing model, different stagesof the processing pipeline overlap in time. More specifically,each graph shows the time difference between the upload starttime and the termination time of the process that belong tothat stage of the pipeline (i.e., the graphs from bottom to topare cumulative).

CONCLUSION AND FUTURE WORK

In order to detect errors in big data sets from sensor networksystems, a novel approach is developed with cloudcomputing. Firstly error classification for big data sets ispresented. Secondly, the correlation between sensor networksystems and the scale-free complex networks areintroduced. According to each error type and the featuresfrom scale-free networks, we have proposed a time-efficientstrategy for detecting and locating errors in big data sets oncloud. With the experiment results from our cloud computingenvironment U-Cloud, it is demonstrated that 1) theproposed scale-free error detecting approach can significantlyreduce the time for fast error detection in numericbig data sets, and 2) the proposed approach achieves similarerror selection ratio to non-scale-free error detectionapproaches. In future, in accordance with error detectionfor big data sets from sensor network systems on cloud, theissues such as error correction, big data cleaning and recoverywill be further explored.A time efficient approach for detecting errors and locating errors in big data sets on cloud through the scale-free networks. By using this proposed approach, we can reduce the time for detecting errors in big data sets. We can overcome the problems of data cleaning and error detection in the bid sensor cloud data by using proposed approach.

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. Under such kind of computing model, more and more enterprises will transfer the storage data to cloud computing center. And the scale becomes more and more large. For now, the research of data storage and management in cloud computing mainly focuses on dealing with data expressing and searching. In this study, we provided a survey on numerous models and approaches of tackling these problems. We review the various approaches and its idea of design. There are still many open issues, such as, mobile data management in cloud environment, backend-support for data-intensive applications of cloud computing, integration of different cloud computing platform, data mining in cloud computing

FUTURE WORK

The future scope of this work in cloud surrounding, preserving privacy for analysis of data, mining and distribution is very challenging issues for research due to huge data sets requires search. For data anonymization investigation of adoption approach to generalization of bottom-up algorithm.

Fault tolerance support toour stream-based approach, an important feature that is alsosupported by the MapReduce-based alternatives. Moreover,we would like to further explore the stream-as-you-go ideain a way to provide a more generally applicable incrementalprocessing framework than an application-specific one.

REFERENCES

[1] A. Ravi Kumar,Dr. M. Ashok, Dr. R. Praveen Sam, "**Error Detection and Cleaning for Big Data Sets from Sensor Network Systems on Cloud",** International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 5, Sep - Oct 2016.

[2] ChinthojuJyothsna, NittalaSwapnaSuhasini, "**An Approach for Detecting and Analyzing Errors of Big Sensor Data on Cloud**", International Journal of Advanced Technology and Innovative Research Volume. 08, IssueNo.10, August-2016, Pages: 1946-1950.

[3] Satoshi Tsuchiya, Yoshinori Sakamoto and Vivan Lee, "**Big Data Processing in Cloud Environemnts"** FUJITSU.Tech. J., Vol.48, No.2, April – 2012.

[4] S. Mukhopadhyay, D. Panigrahi, and S. Dey, **"Model Based Error Correction for Wireless Sensor Networks,"** IEEE Trans. Mobile Computing, vol. 8, no. 4, pp. 528-543, Sept. 2008.

[5] M.C. Vuranand and I.F. Akyildiz, **"Error Control in Wireless Sen-sor Networks: A Cross Layer Analysis,"** IEEE Trans. Networking, vol. 17, no. 4, pp. 1186-1199, Aug. 2009.