IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

Detecting Errors in Big Data on Cloud

M. Pramod hand
MTech(CSE)
SLC's Institute of Engineering and Technology

Adavelli ramesh$_{Mtech}$
Assoc. Prof Dep. of computer Science
SLC's Institute of Engineering and Technology

Vanam gopinath$_{Mtech}$
Assoc. Prof Dep. of computer Science
SLC's Institute of Engineering and Technology

.

Abstract— This Project "DETECTING ERRORS IN BIG DATA ON CLOUD". Big 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. Firstly, a set of sensor data error types are classified and defined. Based on that classification, the network feature of a clustered WSN is introduced and analyzed to support fast error detection and location. Specifically, 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.

Keywords— Bigdata, Cloud Computing, Sensor Networks, Data Abnormality, Error Detection

I. INTRODUCTION

Recently, we enter a new era of data explosion which brings about new challenges for big data processing. In general, big data[1],[2] 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[1], [2], [3], [4], [5], [6]. 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. In this paper, 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. However, 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. The proposed error detection approach in this paper will be based on the classification of error types. Specifically, four types of numerical data abnormalities/errors are listed and introduced in our cloud error detction approach.the main contribution of our proposed detection is to achieve significant time performance improvement in error detection without compromising error detection accuracy.

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

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

detecting accuracy.

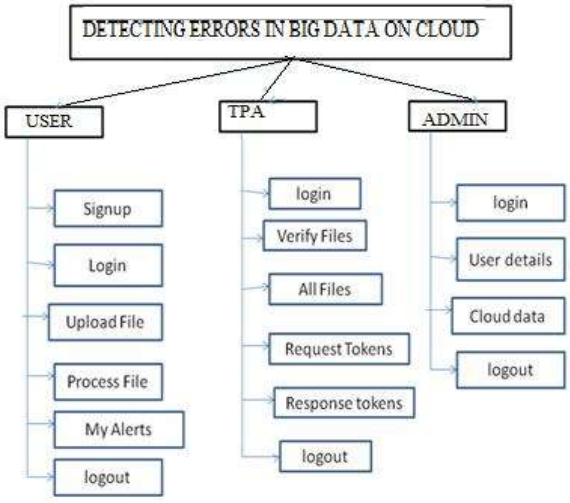
Advantages

- Cloud computing is a disruptive trend at present, poses a significant impact on current IT industry and research communities.
- It provides an open, flexible, scalable and reconfigurable platform.
- Especially, how to use the computation power of cloud to quickly find and locate errors of nodes in WSN needs to be explored.

1. Literature Survey

The limitation of cloud computing are the security issues of cloud computing. It comes to know that here are no security standards available for secure cloud computing. Users have serious concerns about confidential of sensitive cloud. The main security problems involve user data privacy, data security, protection, cloud computing administration and cloud computing platform stability. Customers should have the right of the supervision and have audit of cloud computing services for fully ensure the security of customer data.The data must be protected from virus, worms and Trojan in cloud computing platform within the network of internal and external.

III. SYSTEM ARCHITECTURE



IV. MODULES

In this project we are using 4 modules

1. User module
2. TPA Module
3. Admin module
4. Error definition module

1. User module: In this module the User register to the application with his/her complete details. In this user can upload a file and user can process the file like update, delete or append. User can view all files which they upload.

2. Admin module: In this module Admin will check the details of registered Users. And also admin can check which

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

type of errors will occur in the Data files present in cloud.

3. TPA Module : In this module Third Party Authentication(TPA) will verify the files which are processed by users, view all files which are uploaded by users. For verification purpose TPA request users for tokens

4. Error definition module: In order to test the false positive ratio of our error detection approach and time cost for error findings,we impose 4 types of data errors.

- **Flat Line Fault:** The "flat line faults" indicates a time series of a node in a network system keeps unchanged for unacceptable long time duration.
- **Out of data bounds fault:** The "out of data bounds faults" indicates impossible data values are observed based on some domain knowledge.
- **Data lost fault:** The "data lost fault" means there are missing data values in a time series during the data generation or communication.
- **Spike faults:** The "spike faults" indicates in a time series data items which are totally out of the prediction and normal changing trend.

V. RESULTS

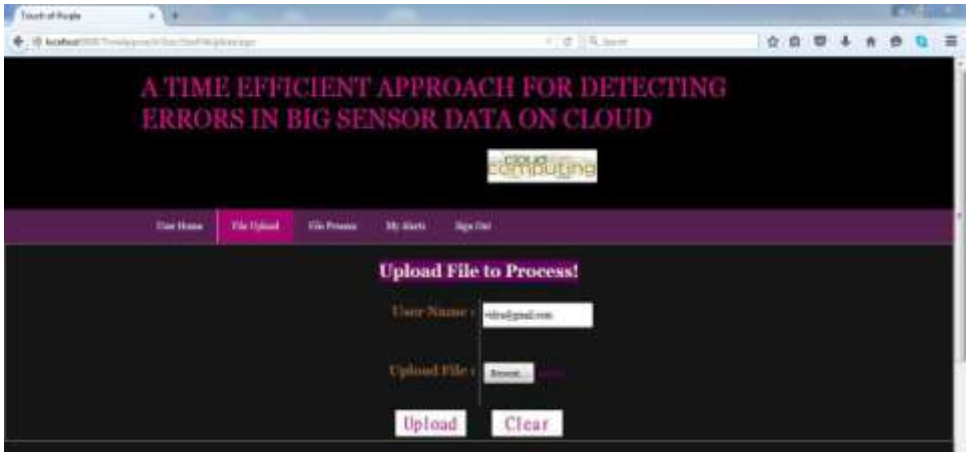


Fig. 5.1: User file upload page

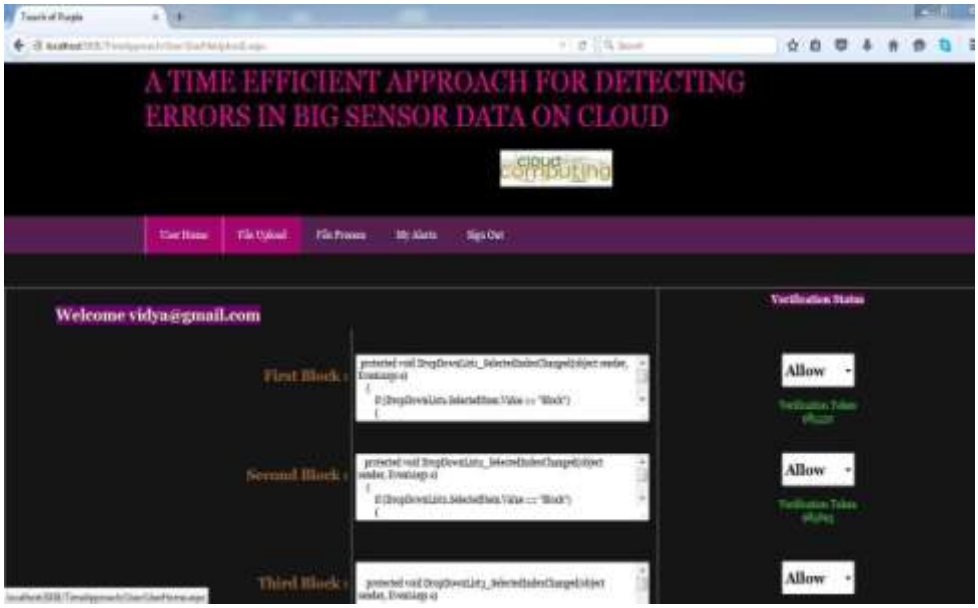


Fig. 5.2: Change file status

Fig. 5.3: Error types page

Fig. 5.4: Errors in file

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

VI. CONCLUSION

In order to detect errors in big data sets from sensor network systems, a novel approach is developed with cloud computing. Firstly error classification for big data sets is presented. Secondly, the correlation between sensor network systems and the scale-free complex networks are introduced. According to each error type and the features from scale-free networks, we have proposed a time-efficient strategy for detecting and locating errors in big data sets on cloud.In future, in accordance with error detection for big data sets from sensor network systems on cloud, the issues such as error correction, big data cleaning and recovery will be further explored.

REFERENCES

[1] S. Tsuchiya, Y. Sakamoto, Y. Tsuchimoto, and V. Lee, "Big Data Processing in Cloud Environments," FUJITSU Science and Technol- ogy J., vol. 48, no. 2, pp. 159-168, 2012.

[2] "Big Data: Science in the Petabyte Era: Community Cleverness Required," Nature, vol. 455, no. 7209, p. 1, 2008.

[3] X. Zhang, C. Liu, S. Nepal, and J. Chen, "An Efficient Quasi-Iden- tifier Index Based Approach for Privacy Preservation over Incre- mental Data Sets on Cloud," J. Computer and System Sciences, vol. 79, pp. 542-555, 2013.

[4] X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen, "A Privacy Leakage Upper-Bound Constraint Based Approach for Cost-effec- tive Privacy Preserving of Intermediate Datasets in Cloud," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 6, pp. 1192-1202, June 2013. [5] X. Zhang, T. Yang, C. Liu, and J. Chen, "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using Systems, in MapReduce on Cloud," IEEE Trans. Parallel and Distributed, vol. 25, no. 2, pp. 363-373, Feb. 2014.

[6] C. Liu, J. Chen, T. Yang, X. Zhang, C. Yang, R. Ranjan, and K. Kotagiri, "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," IEEE Trans. Parallel and Distributed Systems, vol. 25, no. 9, pp. 2234–2244, Sept. 2014.

[7] Chi Yang, Chang Liu, Xuyun Zhang, Surya Nepal, and Jinjun Chen, "A Time Efficient approach for Detecting errors In Sensor Data on Cloud", IEEE Trans. Parallel and Distributed Systems, vol. 26, no. 2, february

[8] S. Mukhopadhyay, D. Panigrahi, and S. Dey, "Data Aware, LowCost Error Correction for Wireless Sensor Networks," Proc. IEEE Wireless Comm. and Networking Conf. (WCNC "04), pp. 2494-2497,2004.

[9] M.H. Lee and Y.H. Choi, "Fault Detection of Wireless Sensor Networks," Computer Comm., vol. 31, no. 14, pp. 3469-3475, 2008.

[10] "Sensor Cloud," http://www.sensorcloud.com/, accessed on 30,Aug. 2013.

[11] K. Ni, N. Ramanathan, M.N.H. Chehade, L. Balzano, S. Nair, S. Zahedi, G. Pottie, M. Hansen, M. Srivastava, and E. Kohler, "Sensor Network Data Fault Types," ACM Trans. Sensor Networks, vol. 5, no. 3, article 25, May 2009.

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363