

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

# Optimized Error Detection Analytics with Bigdata on Cloud

Damini.V.S<sup>1</sup>, Rabindranath<sup>2</sup>, Priyashree.K<sup>3</sup>, Jayashubhaj.K<sup>4</sup>

4<sup>th</sup> Semester M.TECH, Dept. of Computer Science and Engineering, AMC Engineering Collage, Bangalore, India

**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. 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. 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.

## 1. 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], [12], [13], [14], [15], [16], [17], [30], [31], [32], [33]. 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, bio-logical study, gene analysis and environmental research [1], [2]. According to literature [1], [2], 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 10<sup>18</sup>) 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 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 loss-less. However, effective detection and cleaning of sensor big data errors is a challenging issue demanding innovative solutions.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

WSN with cloud can be categorized as a kind of complex network systems [21]. In these complex network systems [21], [22], [23], [24], such as WSN and social network, data abnormality and error become an annoying issue for the real network applications [25], [26], [27]. 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 [28], [30] 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 [37], [38]. Since these techniques were not designed and developed to deal with big data on cloud, they were unable to cope with current dramatic increase of data size. For example, when big data sets are encountered, previous offline methods for error detection and debugging on a single computer may take a long time and lose real time feedback. Because those offline methods are normally based on learning or mining, they often introduce high time cost during the process of data set training and pattern matching. 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. 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 [38], [39]. However, fast detection of data errors in big data with cloud remains challenging.

## II. RELATED WORK AND PROBLEM ANALYSIS

To address various challenges of big data, research works can be found intensively from the database view [30], [31], [32]. However, the problem can be also discussed from the perspective of parallel systems and cloud [35], [36]. In this section, related literature for big data processing on cloud, and data error detection for complex network systems will be reviewed and compared.

### 2.1 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 [5], [6], [7], [8], [9], [10], [11]. 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 [3], [4]. 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 EC2 infrastructure as a service is a typical cloud based distributed system for big data processing. Amazon S3 supports distributed storage. Map Reduce [7], [10], [18], [19], [20] 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. Kienzler et al. [8] designed 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 [11] 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. 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

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

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 the operation.

## 2.2 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 [28], [38], [39] 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 [21], [22], [23], [29]. Sensor-Cloud [27] 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. 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 [37] by Elnahrawy and path. They deal with the problems of outliers, missing information, and noise. A novel online approach for modelling 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 proposed [37]. The proposed approach is efficient in reducing the uncertainty associated with noisy sensors. However, the scalability and error detection accuracy are not dealt. It is an initial and important step for online error detection of WSN. But lots of work still needs to be done. Especially, under the cloud environment, the computational power and scalability should be fully exploited to support the real time fast error detection for sensor data sets. To the best of our knowledge based on the above work, the error detection issues of big data from WSN are rarely discussed on current sensor-cloud technology or other online WSN data processing techniques.

## 2.3 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 [22], [23], [24], [25], [26], [34]. 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. [22] 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 [22] can be extended for the errors in complex network systems. Xiong proposed an approach [23] which can be used to detect the text data

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

errors in data sets of social network. Mukhopadhyay [24] proposed 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 [24] 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 of Mukhopadhyay. Slijepcevic analyses the location errors in sensor networks in [25]. The primary goal of this location error analysis is to demonstrate the practical use of the location errors for optimal resource consumption. Ramanathan presents a detailed study of sensor faults that occur in deployed sensor networks and a systematic approach to model these faults [26]. Sheth develops a decentralized fault diagnosis system for WSN in [28]. It enables efficient management of a WSN by diagnosing the true root cause of a degraded performance by combining multiple sensor observations. This diagnosis requires minimal data collection at the centralized base station. Khan in [29] presents a sensor network troubleshooting tool that helps the developer diagnose root causes of errors. The tool is geared towards finding interaction bugs. Anyway, in the paper [29], it is pointed out that scalability, user interface and detecting time still need to be improved. It can be concluded that current data error detection techniques for complex network systems focus on in-network detecting with intelligent nodes or offline analysis at the root. They ignore the scalability, massive resource and powerful computation capability provided by cloud. The proposed approach in this paper aims to address this issue by utilizing the inherent features of cloud computing to realize fast error detection on cloud.

### III. ERROR TYPES IN WSN BIG DATA SETS

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 [21], [22], [23]. 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.

#### 3.1 Error and Abnormality Classification

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 [22], the errors of complex networks can be classified as six main types for both numeric and text data as Appendix A.1, which can be found on the Computer Society Digital Library.

### IV. TIME-EFFICIENT ERROR DETECTION FOR BIG SENSOR DATA ON CLOUD

According to the above analysis, it is clear that complex network systems have a similar clustered network topology. During the filtering of big data sets, whenever an abnormal data is encountered, the detection algorithm needs to finish two tasks. They are depicted as two functions here. " $f_{dn=e;tp}$ " is a decision making function which determines whether the detected abnormal data is a true error. In other words,  $f_{dn=e;tp}$  has two outputs, "false negative" for detecting a true error and "false positive" for selecting a non-error data. " $f_{dn=e;tp}$ " is a function for tracking and returning the original error source. With the results from the above two functions, the error detection process can be successfully finalized. As shown in Fig. 1, there is a complex network and cloud platform for running error detecting algorithms. Without any consideration of network features and data characteristics, the error detection algorithm needs to filter the whole big data set from the network. Whenever, an abnormality defined in Section 3 is encountered, the algorithm will call  $f_{dn=e;tp}$  and  $f_{dn=e;tp}$  to traverse the whole network big data set

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

for the final decision making and error source location. However, based on the analysis of scale-free network systems, it has been proved that scale-free networks have a clustering and hierarchical topology. Only a few nodes in the whole network have large sets of links to other nodes. So, based on these nodes, the whole networks can be partitioned into a group of clusters (red circles). If there is certain abnormal data occurs for a certain node  $k$ , the high opportunity is that most of the related data for  $f_{d \setminus n=e; t \setminus}$  and  $f_{l \setminus n=e; t \setminus}$  will be located in the clusters where the node  $k$  locates. As a result,  $f_{d \setminus n=e; t \setminus}$  and  $f_{l \setminus n=e; t \setminus}$  only need to navigate the related clusters for error detection result. This is because of the fact that except for a few central nodes, most of nodes only have limited links within themselves in their clusters. Hence, the proposed clustering can significantly reduce the time cost error locating and final decision making by avoiding whole network data processing addition, with this detection technique, cloud resources only need be distributed according to each partitioned cluster in a scale-free complex network.

## V. EXPERIMENTS

To verify the time efficiency and the effectiveness of our approach for detecting errors in big data with cloud, experiments are conducted on U-Cloud (cloud computing environment at the University of Technology Sydney) [12], [13], [14], [15], [16], [18]. There are three purposes for this experiment. 1) Demonstrate that the significant time-saving is achieved in terms of detecting errors from complex network big data sets. 2) Demonstrate the effectiveness of our proposed error detection approach in terms of different error types. 3) Demonstrate that the false positive ratio of our proposed error detection algorithm is limited within a small value.

### 5.1 Experiment Environment and Process

The U-Cloud system is set up as shown in Appendix C.1, available in the online supplemental material. Four types of data values collected by a real WSN (scale-free complex network system) are used as the testing data set. The total testing data set size is around 2,000,000 KB, including temperature, sound, light and vibration. Even only considering one node, four types of testing data are gathered with different frequency. In other words, the data sampling from each real world node is heterogeneous. Before the experiment, we conduct the normalization for the testing data set. The normalization process is described in Appendix C.2, available in the online supplemental material.

### 5.2 Experiment Results

In order to test the false positive ratio of our error detection approach and time cost for error findings, we impose five types of data errors following the definition in Section 3 into the normalized testing data sets with a uniform random distribution. 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 fourth 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.

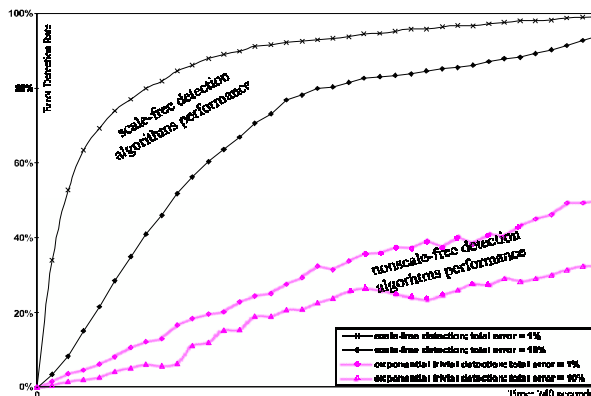


Fig. 1. Time cost for detecting errors from the testing data set

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

In Fig. 1, the testing results show the time performance of our proposed scale-free error detection algorithm on U-Cloud after 740 seconds. Specifically, 10 different error rates are imposed into the experimental data set and tested independently. The testing error rate changes from 1 to 10 percent in 10 repetitive experiments. After about 100 seconds, the proposed algorithm can detect more than 60 percent errors whatever the testing error rate is within the domain between 1 and 10 percent. During the time duration between 0 and 100 second, all error detection rates increase dramatically with a steep trend.

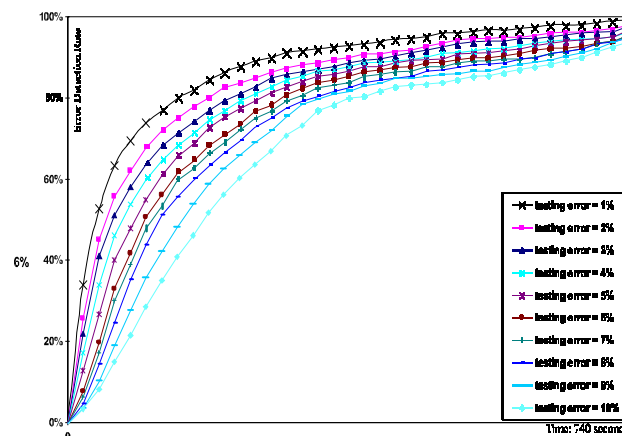


Fig. 2. Comparison of two error detection strategies

## VI. CONCLUSIONS AND FUTURE WORK

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. With the experiment results from our cloud computing environment U-Cloud, it is demonstrated that 1) the proposed scale-free error detecting approach can significantly reduce the time for fast error detection in numeric big data sets, and 2) the proposed approach achieves similar error selection ratio to non-scale-free error detection approaches. 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 Technology 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] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwin-ski, 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.
- [4] 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.
- [5] 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. 2012.
- [6] S. Sakr, A. Liu, D. Batista, and M. Alomari, "A Survey of Large Scale Data Management Approaches in Cloud Environments," IEEE Comm. Surveys & Tutorials, vol. 13, no. 3, pp. 311-336, Third Quarter 2011.
- [7] B. Li, E. Mazur, Y. Diao, A. McGregor, and P. Shenoy, "A Platform for Scalable One-Pass Analytics Using MapReduce," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD'11), pp. 985-996, 2011.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

- [8] R. Kienzler, R. Bruggmann, A. Ranganathan, and N. Tatbul, "Stream As You Go: The Case for Incremental Data Access and Processing in the Cloud," Proc. IEEE ICDE Int'l Workshop Data Management in the Cloud (DMC'12), 2012.
- [9] C. Olston, G. Chiou, L. Chitnis, F. Liu, Y. Han, M. Larsson, A. Neumann, V.B.N. Rao, V. Sankarasubramanian, S. Seth, C. Tian, T. ZiCornell, and X. Wang, "Nova: Continuous Pig/Hadoop Workflows," Proc. the ACM SIGMOD Int'l Conf. Management of Data (SIGMOD'11), pp. 1081-1090, 2011.
- [10] K.H. Lee, Y.J. Lee, H. Choi, Y.D. Chung, and B. Moon, "Parallel Data Processing with MapReduce: A Survey," ACM SIGMOD Record, vol. 40, no. 4, pp. 11-20, 2012.
- [11] "Hadoop," <http://hadoop.apache.org>, accessed on March 01, 2013.
- [12] X. Zhang, C. Liu, S. Nepal, and J. Chen, "An Efficient Quasi-Identifiable Index Based Approach for Privacy Preservation over Incremental Data Sets on Cloud," J. Computer and System Sciences, vol. 79, pp. 542-555, 2013.
- [13] X. Zhang, C. Liu, S. Nepal, S. Pandey, and J. Chen, "A Privacy Leakage Upper-Bound Constraint Based Approach for Cost-effective Privacy Preserving of Intermediate Datasets in Cloud," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 6, pp. 1192-1202, June 2013.
- [14] 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.
- [15] 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.
- [16] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications," IEEE Trans. Parallel and Distributed Systems, 2013.
- [17] C. Yang, X. Zhang, C. Zhong, C. Liu, J. Pei, K. Kotagiri, and J. Chen, "A spatiotemporal compression based approach for efficient big data processing on cloud," J. Computer and System Sciences, vol. 80, no. 8, pp. 1563-1583, 2014.
- [18] J. Conhen, "Graph Twiddling in a MapReduce World," IEEE Computing in Science & Eng., vol. 11, no. 4, pp. 29-41, 2009.
- [19] K. Shim, "MapReduce Algorithms for Big Data Analysis," Proc. VLDB Endowment, vol. 5, no. 12, pp. 2016-2017, 2012.
- [20] "Big Data Beyond MapReduce: Google's Big Data Papers," <http://architects.dzone.com/articles/big-data-beyond-mapreduce>, accessed Mar. 2013.
- [21] R. Albert, H. Jeong, and A. L. Barabasi, "Error and Attack Tolerance of Complex Networks," Nature, vol. 406, pp. 378-382, July 2000.
- [22] D.J. Wang, X. Shi, D.A. McFarland, and J. Leskovec, "Measurement Error in Network Data: A Re-Classification," Social Networks, vol. 34, no. 4, pp. 396-409, Oct. 2012.
- [23] D. Xiong, M. Zhang, and H. Li, "Error Detection for Statistical Machine Translation Using Linguistic Features," Proc. 48th Ann. Meeting of the Association for Computational Linguistics (ACL'10), pp. 604-611, 2010.
- [24] 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.
- [25] S. Slijepcevic, S. Megerian, and M. Potkonjak, "Characterization of Location Error in Wireless Sensor Networks: Analysis and Application," Proc. the Second Int'l Conf. Information Processing in Sensor Networks (IPSN '03), pp. 593-608, 2003.
- [26] 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.
- [27] A. Alamri, W.S. Ansari, M.M. Hassan, M.S. Hossain, A. Alelaiwi, and M.A. Hossain, "A Survey on Sensor-Cloud: Architecture, Applications, and Approaches," Int'l J. Distributed Sensor Networks, vol. 2013, pp. 1-18, 2013.
- [28] A. Sheth, C. Hartung, and Richard Han, "A Decentralized Fault Diagnosis System for Wireless Sensor Networks," Proc. IEEE Second Conf. Mobile Ad-hoc and Sensor Systems (MASS '05), Nov. 2005.
- [29] M.M.H. Khan, H.H.K. Le, H. Ahmadi, T.F. Abdelzaher, and J. Han, "Dustminer: Troubleshooting Interactive Complexity Bugs in Sensor Networks," Proc. ACM Sixth Conf. Embedded Network Sensor Systems (SenSys '08), pp. 99-112, 2008.
- [30] N. Laptsev, K. Zeng, and C. Zaniolo, "Very Fast Estimation for Result and Accuracy of Big Data Analytics: The EARL System," Proc. IEEE 29th Int'l Conf. Data Eng. (ICDE), pp. 1296-1299, 2013.
- [31] X.L. Dong and D. Srivastava, "Big data integration," Proc. IEEE 29th Int'l Conf. Data Eng. (ICDE), pp. 1245-1248, 2013.
- [32] T. Condie, P. Mineiro, N. Polyzotis, and M. Weimer, "Machine learning on Big Data," Proc. IEEE 29th Int'l Conf. Data Eng. (ICDE), pp. 1242-1244, 2013.
- [33] A. Aboulnaga and S. Babu, "Workload Management for Big Data Analytics," Proc. IEEE 29th Int'l Conf. Data Eng. (ICDE), p. 1249, 2013.
- [34] S. Mukhopadhyay, D. Panigrahi, and S. Dey, "Data Aware, Low Cost Error Correction for Wireless Sensor Networks," Proc. IEEE Wireless Comm. and Networking Conf. (WCNC '04), pp. 2494-2497, 2004.
- [35] M.H. Lee and Y.H. Choi, "Fault Detection of Wireless Sensor Networks," Computer Comm., vol. 31, no. 14, pp. 3469-3475, 2008.
- [36] M.C. Vuran and I.F. Akyildiz, "Error Control in Wireless Sensor Networks: A Cross Layer Analysis," IEEE Trans. Networking, vol. 17, no. 4, pp. 1186-1199, Aug. 2009.
- [37] E. Elnahrawy and B. Nath, "Online Data Cleaning in Wireless Sensor Networks," Proc. First Int'l Conf. Embedded Networked Sensor Systems (ACM Sensys '03), pp. 294-295, 2003.
- [38] M. Yuriyama and T. Kushida, "Sensor Cloud Infrastructure," Proc. 13th Int'l Conf. Network-Based Information Systems (NBIS), pp. 1-8, 2010.
- [39] "SensorCloud," <http://www.sensorcloud.com/>, accessed on 30, Aug. 2013.