2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)
Dec 20-22, 2013, Shenyang, China

An Integrated Framework for Managing Massive and Heterogeneous Sensor Data Using Cloud Computing

Xin Song , Cuirong Wang, Yanjun Chen
Computing Center
Northeastern University at Qinhuangdao
Qinhuangdao, China
bravesong@163.com

Abstract—**Cloud computing can provide a powerful, scalable storage and the massive data processing infrastructure to perform both online and offline analysis and mining of the heterogeneous sensor data streams. With the recent explosion of wireless sensor networks and their applicability in military and civilian applications, there is an emerging vision for integrating sensor networks into the cloud computing platform. In contrast to traditional data objects, the sensor data objects have continuously changed, high-dimensional, spatiotemporal relation and heterogeneous attributes. Therefore, the management and processing problem of the massive sensor data objects can be more complicated. The paper formally presents an integrated framework for managing massive and heterogeneous sensor data with insights into the high-dimensional problem using the map-reduce platform of cloud computing. The proposed framework incorporates key concepts such as parallel-processing, scalability and flexibility of resources, sensor data uncertainty and the dynamic deployment and management of applications.**

Keywords—Massive sensor data management; Cloud computing; Parallel processing; MapReduce

I. INTRODUCTION

In recent years, an increasing number of emerging applications deal with a large number of heterogeneous sensor data objects in Internet of Things (IoT) because of a wide variety of sensor devices on sensing layer. The large-scale heterogeneous sensor stream processing becomes the key issue for the IoT system application. In order to process the continuously changing sensor data streams, the IoT application system terminal equipment must implement the massive sensor data storage and the powerful computing ability for real-time collection, dissemination and extracting of sensor data to users and administrators anytime and anywhere. However, due to the continuity and infiniteness of the sensor data streams, the traditional sensor data processing model focuses on a relatively small scale historical data and is not able to meet the increasing data processing requirement of IoT system because of some restricted factors such as memory capacity, data collection speed, transmission bandwidth and so on. The most important and interesting research issues to developing a massive and heterogeneous sensor data management system are related to storing data trajectories and processing real-time inquire services. Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources [1]. With the mass data processing ability and storage resources of cloud computing, the IoT system will greatly improve the efficiency of operation to provide timely and effective sensor information services.

In this paper, we propose an integrated framework for managing massive and heterogeneous sensor data with insights into the high-dimensional problem using the map-reduce platform of cloud computing. The proposed framework incorporates key concepts such as parallel-processing, scalability and flexibility of resources, sensor data uncertainty and the dynamic deployment and management of applications.

The remainder of the paper is organized as follow: In Section 2, we describe and discuss some of the related work about the sensor data management using cloud computing technology. The proposed framework setup and analytical results are discussed in Section 3. Finally, we conclude the paper in Section 4.

II. RELATED WORK

There have been a few of studies on the management of sensor data using cloud computing. More recently developed large-scale sensory data processing system is presented in Ref. [2]. It proposed the idea of distributed databases to store sensory data and Map Reduce programming model for large-scale sensory data parallel processing. The system improved the data processing capability of WSNs, but the multiple models, multiple attributes and high-dimensional sensor data processing technologies for the internet of things are not discussed. Doukas Charalampos and Maglogiannis Ilias developed and present a wearable-textile platform based on open hardware and software that collects motion and heartbeat data and stores them wirelessly on an open Cloud infrastructure for monitoring and further processing. The proposed system may be used to promote the independent living of patient and elderly requiring constant surveillance [3]. Ref. [4] specifically targets at extracting useful information from the accelerometer sensor data. The paper proposed the utilization of parallel computing using MapReduce on the cloud for training and recognizing human activities based on classifiers that can easily scale in performance and accuracy. The sensor data is extracted from the mobile, offloaded to the cloud computing platform and processed using three different

The research work was supported by the Fundamental Research Funds for the Central University under Grant no. N120323009, and the Doctoral Fund of Northeastern University at Qinhuangdao under Grant no. XNB201301.

classification algorithms. Moreover, Ref. [5] and Ref [6] focused on the real-time health monitoring system using cloud computing technology. The important ideas of these research works are proposing a novel architectural design and a use case for integrating cloud and mobile computing technologies to realize a health monitoring and analysis system. In order to combine WSN and cloud computing for implementing the performance improvement of on-demand sensing service in Internet of things, in the following different proposals are analyzed. The distributed shared memory concept of reflecting WSNs inside the cloud by virtual sensors was introduced in Ref [7]. The paper presented a middleware named tinyDSM which provides means for replicating sensor data and ensuring the consistency of the replicates. Because each application manages both of physical sensors and sensor data, users, having no specialty physical sensors, cannot use the specialty monitoring system directly. There is a sensor-cloud infrastructure on which users could share multiple different kinds of physical sensors easily. It provided virtual sensors so the users need not worry about the real locations and differences of multiple physical sensors. The users can use and control virtual sensors with standard functions [8-11]. The objective of an integration framework between cloud computing and WSNs is to facilitate the shift of sensor data from WSNs to the cloud computing environment so that the scientifically and economically valuable data may be fully utilized [12-15]. An increasing number of data-intensive application deals with continuously changing data streams from sensors. One requires the data processing system that can store, update, and retrieve large sets of multidimensional sensor data. The conventional information system technology cannot manage the continuously changing properties of the sensor data [16]. Therefore, it is very necessary for managing massive and heterogeneous sensor data via combining the cloud computing and WSNs technology.

III. IMPLEMENTATION OF INTEGRATED FRAMEWORK FOR MANAGING MASSIVE AND HETEROGENEOUS SENSOR DATA USING CLOUD COMPUTING

To build a large-scale heterogeneous sensing data stream processing system based on cloud computing, the managing platform must have both the distributed storage technology of massive sensor data with high speed data flow characteristics and real-time response algorithm with effectively reducing data processing delay. Furthermore the highly effective mass heterogeneous sensor data distributed knowledge discovery and the low complexity parallel data mining algorithm are equally important.

A. Distributed Storage Solution of the Polymorphism Sensor Data Streams in Cloud Computing

The sensor nodes in the sensing layer of Internet of Things are different variety, including some environment sense devices, video monitoring terminal, RFID tag and so on. The data structure of each device is not identical. So it is necessary to design a distributed storage solution of the polymorphism sensor data stream in cloud computing according to the characteristics of the IoT system application. In the proposed framework, the Distributed storage solution of the polymorphism sensor data stream in cloud computing environment is three-level storage architecture, as shown in figure 1. The operation support data layer is responsible for the storage and dynamic update of the sensor data streams (or the intermediate results). The operation result data layer is responsible for the storage and dynamic update of the final processing results. The historical data layer is responsible for the storage and additional update of the historical sensor data. The candidate historical data in the operation result data layer were stripped out and appended to the historical data layer after each time data processing. The central storage scheduling module controls respectively the three layers data collection according to the relevant instruction and keeps the data consistency between the operation support data layer and the operation result data layer in the system operation process. The basic storage management layer provides the data acquisition and update services using the cluster distributed file system.

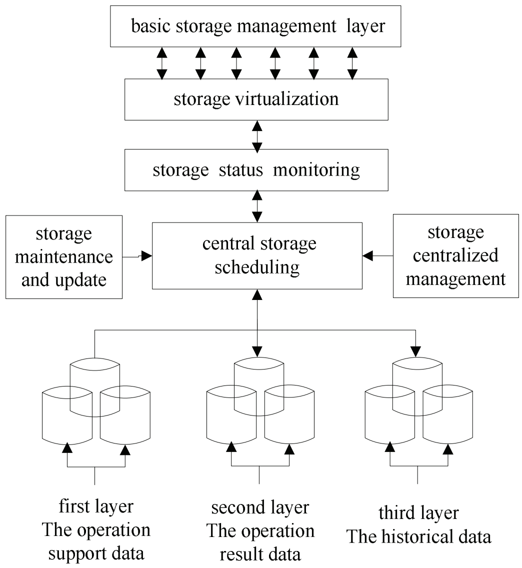

Fig. 1. The three-level storage architecture

B. The Map-Reduce Processing Strategy of Massive Sensor Data Streams

The sensor data status may be rapidly changing whether in the WSN or the RFID system. The sensor data acquisition was operated at any time. Unlimited historical data backup not only consume a large amount of storage space but also affect the efficiency of data processing and query. Therefore, the proposed framework must reduce the latency of data processing and improve the reliability of the system on the basis of the sensor data streams characteristics preserving. To ensure real-time processing of a large-scale heterogeneous sensor data streams in cloud computing, the intermediate results of the preprocessing historical data were distributed at each cache nodes for reducing the duplication processing of the historical sensor data and avoiding the frequent transmission between nodes. Each node redundantly received data stream, so the pending processing data of the node were filtered by Map stage and operated Reduce calculation in the node local cache. When the existing node's local computing and storage resources cannot meet the real needs, the new increased node will be utilized for mobile cache data extension using the re-division technology. Finally, the local calculation results were synchronized to the distributed storage area. The Map-Reduce process is shown in figure 2.

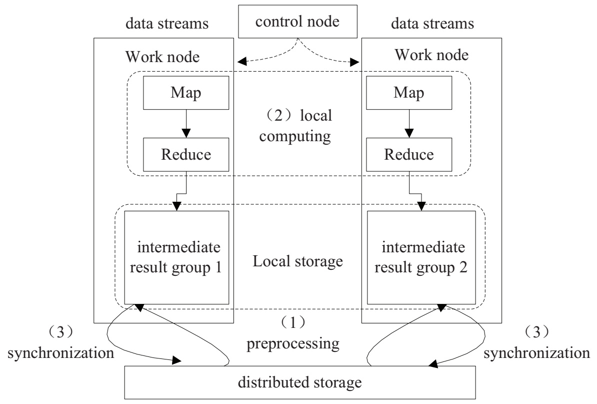

Fig. 2. The workflow of the sensor data streams processing

C. *The parallel processing scheme of the massive sensor data streams for analysis of the interested information*

Most of the raw data in the IoT system is unstructured, for example, figure structure, sequence and the continuous measurements, etc. The traditional data mining algorithms cannot be directly applied in cloud computing for the sensor data stream that cannot use the feature vector representation. The proposed framework realized the cloud resources independent dynamic allocation and scheduling for the massive sensor data mining using virtualization and unstructured data extracting technologies. The traditional parallel data mining algorithm is to assume that all pending processing data once loaded into memory. It is not suitable for processing the huge amounts of heterogeneous sensor data sets. The parallel processing scheme of the massive sensor data streams for analysis and mining of the interested information can be used as shown in figure 3.

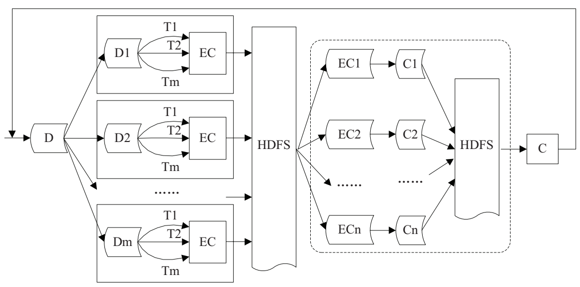

Fig. 3. The parallel processing scheme of the massive sensor data streams

The large-scale sensor data is divided into the multiple data sharding, and then the exported equivalence class of the different candidate properties set is computing using parallel processing strategy on each data sharding. The total number of the recognizable or unrecognizable object pair can be obtained using the data parallel computing scheme after the large-scale recognizable or unrecognizable object pair of the candidate properties is generated. Finally, the best candidate properties are determined.

IV. CONCLUSION

The wireless sensor networks are used to sense and collect environmental information. However, the WSNs are limited in their processing power, storage capacity, battery life and communication speed and so on. Cloud computer technology usually provides the necessary storage capacity and processing power for the massive data analysis and extracting. In this paper, an integrated framework was proposed for managing massive and heterogeneous sensor data using the map-reduce platform of cloud computing. The proposed framework solved three important issues, at first the massive data stream storage scheme was implemented by the three-level storage architecture. Furthermore, the Map-Reduce Processing Strategy of Massive Sensor Data Streams was presented for improving the real-time processing performance. Finally, the proposed framework realized the cloud resources independent dynamic allocation and scheduling for the massive sensor data mining using parallel processing and unstructured data extracting technologies.

However, there are some open issues in combining WSN and cloud computing. for example, the cloud security issue referring to virtualization, cloud storage and service models. The future work should investigate the security issues for the proposed framework and implement the classification and dimension reducing of the massive sensor data streams for improving the efficiency and accuracy of data processing.

REFERENCES

[1] P. McDaniel, S. W. Smith, “Outlook: cloud computing with a chance of security challenges and improvements”, in: Proceeding of the IEEE computer and reliability societies, pp. 77-80, 2010

[2] B. Tang, Y. Wang, “Design of large-scale sensory data processing system based on cloud computing”, Research Journal of Applied Sciences, Engineering and Technology. 4(8),pp. 004-1009, 2012

[3] C. Doukas and I. Maglogiannis, “Managing wearable sensor data through cloud computing”, In: Proceedings of 3rd IEEE International Conference on Cloud Computing Technology and Science (CloudCom) , pp. 440-445. IEEE Press, Athens, Greece , 2011

[4] P. Carlos, F. Huber, N. S. Satish, “Mobile sensor data classification for human activity recognition using MapReduce on cloud”, Procedia Computer Science, 10, pp.585-592, 2012

[5] F. Giancarlo, P. Mukaddim and D. F. Giuseppe, “BodyCloud: integration of cloud computing and body sensor networks”, In: Proceeding of theIEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), pp.851-856. IEEE Press, Taipei, Taiwan, 2012

[6] S.Pandey,W.Voorsluys,S.Niu,etal.“An autonomic cloud environm-ent for hosting ECG data analysis services”, Future Generation Computer Systems. 28(1), pp.147-154, 2012

[7] P. Langendoerfer, K. Piotrowski, M. Díaz, et al. “Distributed shared memory as an approach for integrating WSNs and cloud computing”, In: Proceeding of 5th International Conference on New Technologies, Mobility and Security. Istanbul, Turkey,2012

[8] A. Alamri, W.S. Ansari, M.M. Hassan, et al. "A survey on sensor-cloud: architecture, applications, and approaches" International Journal of Distributed Sensor Networks, 2013

[9] N. Poolsappasit,V.Kumar,S. Madria, S. Chellappan, "Challenges in secure sensor-cloud computing" Secure Data Management. In: Proceedings of 8th VLDB Workshop (SDM), pp. 70-84, Seattle, WA, USA, 2011

[10] M. Yuriyama,T. Kushida, "Sensor-cloud infrastructure physical sensor management with virtualized sensors on cloud computing", In: Proceeding of 13th International Conference on Network-Based Information Systems, pp. 1-8, Gifu, Japan ,2010

[11] J. Melchor, M. Fukuda, "A design of flexible data channels for sensor-cloud integration", In: Proceeding of 21st International Conference on Systems Engineering, pp. 251-256, Las Vegas, NV, USA, 2011

[12] K. Ahmed, M. Gregory, "Integrating wireless sensor networks with cloud computing" In: Proceedings of the Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN), pp. 364-366, Beijing, China, 2011

[13] W. Kurschl, W. Beer, "Combining cloud computing and wireless sensor networks" In: Proceeding of the 11th International Conference on Information Integration and Web-based Applications and Services, pp. 512-518, Kuala Lumpur, Malaysia, 2009

[14] P. You, Y. Peng, Z. Li, "An integration framework of cloud computing with wireless sensor networks", Lecture Notes in Electrical Engineering, vol. 214, pp. 381-387, 2013

[15] S. H. Shah, F. K. Khan, W. Ali, J. Khan, "A new framework to integrate wireless sensor networks with cloud computing" In: Proceeding of IEEE Aerospace Conference, Big Sky, MT, United states, 2013

[16] Y. Byunggu, S. Ranjan, H.J. Dong, "An integrated framework for managing sensor data uncertainty using cloud computing", Information Systems. doi: 10. 1016 / j. is. 2011. 12. 003 (2012)