

Application and Implementation of Private Cloud in Agriculture Sensory Data Platform

Shuwen Jiang^{1,2,3,4}, Tian'en Chen^{1,2,3,4(✉)}, and Jing Dong^{1,2,3,4}

¹ Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China

² National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

{jiangsw, chente, dongj}@nercita.org.cn

³ Key Laboratory of Agri-Infomatics, Ministry of Agriculture, Beijing 100097, China

⁴ Beijing Engineering Research Center of Agricultural Internet of Things, Beijing 100097, China

Abstract. With the explosive development of the Internet of things technology in recent years, the Internet of things technology is also used more and more widely in modern agricultural production. For mass sensor data was produced by the Internet of things in agricultural production, While big data bring many benefits and unprecedented challenges to users. The Internet of things in agriculture production produces some complexity problem which are mass sensor data's Scale, sensor data's heterogeneity and mass sensor data's operation, distribution of sensor, high concurrency of data is written etc. In the presence of these problems, this paper put forward a kind solution of agricultural private cloud sensor data Platform, which is named "Sensor PrivateClouds Platform" (SPCP). The Private cloud platform including following modules, All of these are distributed sensor data caching module based on cluster of memcached and Nginx load (SensorCache); heterogeneous data adapter of sensor module (SensorAdpter), distributed computing storage module based on hadoop' HDFS (SensorStorage), efficient query module of sensor data warehouse based on the Hive (SensorStore), management module of sensor metadata (SensorManager), parallel sensor data analysis module (SensorNum) based on the map-reduce of the hadoop, cloud service of sensor data module (SensorPublish) based on webservice. The experimental results show that SPCP have had the abilities which are mass sensor data storage, cleaning of heterogeneous sensor data, real-time query and processing of mass sensor data. These abilities provides a feasible solution for the heterogeneous data storage and mass sensor data's query in the Internet of things of agriculture production.

Keywords: Internet of things · Private cloud · Big data · Hadoop · HDFS

1 Introduction

For improving the efficiency of agricultural production, the Internet of things technology also has been widely used in modern agriculture, its main purpose is to collect plenty of sensor data in agricultural production, through the analysis of the sensor data for the

agricultural production and utilization to further improve the efficiency of agricultural production. However, in the face of characteristic that are wide area of the production, poor environmental conditions, weak signal of communications in agricultural production, collection and utilization of the sensor data is faced with many challenges. For IOT to compared with Internet, facing the core issues is the storage and query of mass heterogeneous sensor data, processing a large number of sensor's intelligent analysis and work together, complex events such as automatic detection and effective coping [1]. These technology research of problems are relatively limited. Through the application of sensor data in agricultural production, agricultural Internet of things can be found the following problems in the sensory data is:

The big sensor data. Internet of things of agriculture usually contain vast amounts of sensors nodes. Most of these sensors such as GPS sensors, temperature sensors, humidity sensors, these sensors are usually deployed in many different parks or more agricultural production environment in the form of group sensor network [2]. And each sensor can timing acquire the latest sensor data, the sensor data gathering to the storage of sensor data network node. The storage nodes is more than store the recent sensor data and in most cases also need to store a historical data such as 1 year all the historical sensor data value for meeting the needs of the complicated data's processing and analysis. As you can imagine the above data is huge, ordinary server storage for huge amounts of sensory data storage, transmission, query and analysis will be an unprecedented challenge. Heterogeneity of sensory data. The sensor data collection network nodes may include different kinds of sensors such as the environment parameter sensors, geographic information sensor, geological, meteorological sensors, video sensor and so on in a large agricultural production environment. While each kind of sensor also includes many specific sensors such as environment parameters of sensors can be subdivided into soil temperature sensor, a soil moisture sensor, CO_2 sensor, light intensity sensor, air temperature sensor, humidity sensor and oxygen sensors. These sensors are not only the structure and function is different, and the format of the sensor data according to the design of the sensor are also different. This will cause the heterogeneity of sensory data [3]. This heterogeneity has greatly enhanced the difficulty of software development and data processing. According to the above problem, this paper proposes a agricultural sensor data Platform based on private cloud which named "Sensor PrivateClouds Platform" (SPCP).

2 Overall Design of Private Cloud in Agricultural Sensor Data Platform

SPCP platform mainly includes the design of Sensor-Cache, which is deploy on a cluster server of MemerCached and Nginx, while is designed as a distributed cache component; Sensor-Adppter which is a adapte component of multi-source heterogeneous sensor data; Sensor-Storage is designed as a distributed storage server based on Hadoop Distributed File System, Sensor-Store is a kind of warehouse which contains sensor meta data, Sensor-Store can efficiently query and compute sensor data via the Hive; Sensor-Manager realize a management system of sensor basic information; Sensor-Num is a

component of distributed computing based on the map-reduce of the hadoop; Sensor-Publish is designed as a Webservice which realize sensor data’s service.

Overall architecture of the SPCP is four levels, First Level is interface level, It contain SensorCache and SensorAdppter. First level mainly complete sensor data’s receiving and transforming. Second level is Applications level. It contains SensorPublish and SensorNum. Second level mainly provide private cloud service of analysis and publish. Third Level is manage level which contains SensorManager, it mainly complete management of sensor data and cloud module. Fourth level is storage level which contains SensorStorage and SensorStore. It mainly complete storage of sensor data in private cloud platform. SensorStorage and SensorStore mainly implements the data storage, operation, and query of huge mass data [4]. The platform used the hadoop distributed file system, in order to HBase as non-relational data storage, HIVE for processing and query of mass sensor data. Building sensor data management system on top of hadoop platform structures, the user can manage data access of sensor data node, real-time monitoring, history query and other functions through the web mode. Overall architecture of sensor data network platform is shown below (Fig. 1).

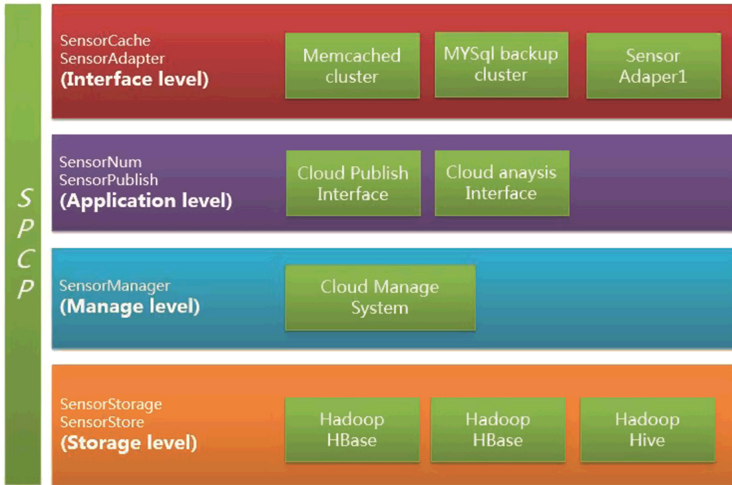


Fig. 1. Overall architecture of sensor data web platform

3 The Design of SensorCache Based on Private Cloud in Hadoop Sensor Data Platform

Sensor data platform use the hadoop distributed file system for data storage and query, now the hadoop platforms use HBase column storage methods for storage, use Hive data warehouse for data query operation, but the experiment proved that real-time latency situation is obvious when the currently used Hive query mass data. It is the problems the entire sensor data platform is facing at present. According to the above problem [5], this paper studies a kind of the cache cluster architecture on top of sensor data platform,

cache cluster cache sensor data for a period of time and use the Mysql relational database for sensor data’s disaster backup in a period of time. This cache cluster can achieve real-time effect in mass data query. The structure of the cache cluster diagram is shown below (Fig. 2).

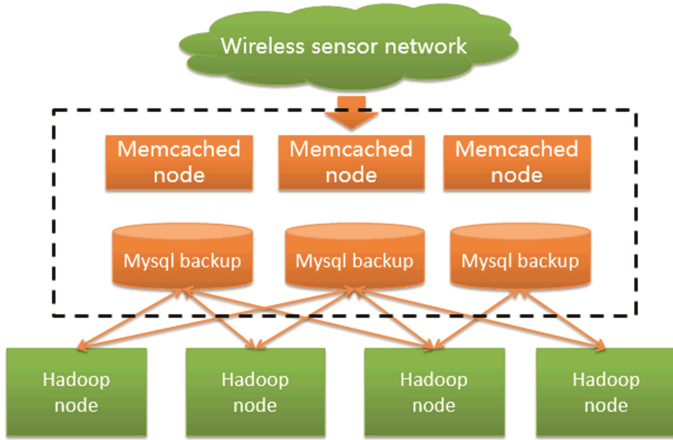


Fig. 2. The overall architecture of the cache cluster diagram

3.1 The Design of the Memcached and Mysql Cluster

The design of Memcached and Mysql cluster is a kind of method based on querying optimization and sensor data backup of Memcached and Mysql cluster. The cluster is used Memcached cluster as sensor data cache service that does not have direct access sensor data from Mysql and liberating the real database. The Mysql so only is used for the history the sensor data’s backup, so as to reduce load when the hadoop query mass data. Cluster’s cache is make full use of multiple servers network memory, CPU and server memory, its basic process of data access as shown in Fig. 1, When the cache cluster is started, the cache cluster first query out the all sensor data for one day from hadoop data platform by using Hive, and stored the sensor data in the corresponding Memcached server through the distributed algorithm, so that in the next trip, user can obtained the sensor data from the corresponding Memcached server. When number of sensor data reach the specified value, the cache server automatically delete useless cache based on the LRU algorithm, the frequency of accessing database will have obvious drop even zero, it is good for mass data query service. Even database mainly perform sensor data disaster backup, the low configuration database also can easily complete [6].

Sensor data’s high concurrent writing is divided into three steps, the first step happens in the cache strategy layer, sensor data will be sent by wireless sensor networks into Memcached nodes via the first layer of the load polling, Memcached node will cache the sensor data as gateway node for the key, sensor data for the value. The second is concurrent sensor data archived Mysql is used to backup history data. Finally, in the time of few users, the cache cluster enable the cache data write to hadoop’s sensor data

service platform in a distributed file system. The cache cluster meet mass sensor data storage, also has reached the efficient querying requirement of real-time mass sensor data.

4 The Design of SensorManager Based on Private Cloud in Hadoop Sensor Data Platform

Sensor data management system is deployed on a sensor data cluster as a management system. It was mainly used in data access of the wireless sensor network, management of nodes, real-time monitoring of sensor data, query of historical sensor data and simple mass sensor data processing, etc.

Sensor data management system is build with J2EE enterprise architecture. The server used the Spring MVC framework, front end used jquery + backbone. Js + HTML5. Sensor data management system is set up in the hadoop data platform between hadoop cluster and cache strategy cluster, cache cluster mainly complete the sensor data query service as query of real-time and query of short-term history sensor data. Query and analysis of mass sensor data are processed in the hadoop distributed file system [7].

Sensor data management system provides the user for a kind of cloud services of sensor data as entrance door of private cloud. The graph is sensor data management system’s operating interface (Fig. 3).



Fig. 3. Data management system of hadoop sensor data platform.

5 The Design of SensorStorage and SensorStor in Sensor Data Platform

Agricultural sensor data’s structure is single and mass. Using relational database for persistence in the early stages that does not take into the characteristics of sensor data’s large scale and distributed. With the development of IOT technology, relational database

processed data and have a bottleneck. The general solution is that copy tables or distribute storages on different server's partition, but the cost of installation and operation is very high. Through distributed storage system(such as HDFS) technology, the server can dynamically change storage nodes by elasticity features, while the existing storage way of sensor data will not be changed sensor data is distributed on server cluster [8].

Hadoop sensor data storage platform is designed by hadoop open source framework which is used by the hadoop's HDFS system. The platform used of natural database HBase in hadoop, and make a distributed storage of mass sensor data. Its advantage lies in the dynamic increase and decrease of distributed cluster, backuping redundancy and efficient distributed computing.

The sensor data platform adopts 6 server cluster, including one namenode, four data nodes, one manager zookeeper. Sensor data platform obtained history sensor data from the cache cluster and transfer to HBase. Using simple analysis service interface of mass data via the HIVE. Figure 4 for architecture diagram of hadoop sensor data platform.

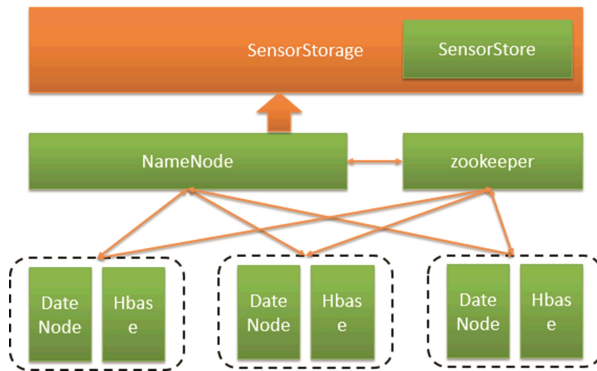


Fig. 4. Architecture diagram of hadoop sensor data platform.

5.1 The Design of NoSql's Sensor Data

Sensor data's storage contains two NoSql relationship tables, the tables respectively contain the table of sensor's meta information and table of sensors original data. The table of sensor's meta information is mainly designed to save sensor's meta basic information. Among the table, SensorId is designed for a column family as a row with the field named sensorInfo. SensorInfo store sensor's meta information by key-value pairs such as sensor-name, sensor-coordinates, sensor-info for a column. Which contains (sensorInfo: id, sensorInfo: net-IP, sensorInfo: net-port, sensorInfo: region servers, sensorInfo: avaCapacity, sensorInfo: location). The table of sensors' original data mainly store original sensors' data of sensor. The table has a raw which contains sensorId and reverse timestamp, field "data" is designed for a family of column which contains(data: temperature, data: carbon-dioxide, data: soil-temperature, data: humidity, data: soil-humidity, data: light) [10].

How big data of sensors can be efficiently read is the key technology in DSM platform. A primary key of “sensorid” is used for a row’s identification which describe the table of sensor information. While row key used a id which combines sensor’s identification and timestamp for the table of sensor original data. So that we can query real-time data from sensor original data via timestamp. As shown in the Table 1.

Table 1. Design of sensor data table

RowKey	times tamp	timeColumnFamily : sensordata			
sensorGroupid	time1	sensordata:tem	sensordata:hum	sensordata:sun	sensordata:co2
	time2	temValue1	humValue1	sunValue1	co2Value1
	time3	temValue2	humValue2	sunValue2	co2Value2

6 Conclusions

This paper analyzes the problems of big sensor data’s acquisition, mass storage and big data analysis in the background of development and application of the IOT and cloud computing technology, with the IOT and cloud computing were wildly applied in the modern agricultural production, these problems brought about the bottleneck is worthy of our attention and solve. To solve the problems existing in the storage of big data, this paper designed a architecture of SPCP, build a wireless sensor network clustering, cache policy cluster and sensor data storage based on hadoop platform. Wireless sensor network realize sensor data acquisition in high concurrency and load balancing, the cache cluster realize efficient real-time query and data backup, data storage platform based on hadoop has realized the huge mass data distributed storage and parallel computing. Experiments show that using SPCP is good at accessing concurrent data, high efficient query and calculation, mass data storage, solved a series of problems brought by the big data.

Acknowledgements. The research work was supported in part by Beijing Natural Science Foundation of China (key project) - Research of adaptive the key technology and application model in agricultural cloud service (4151001) and The Beijing science and technology plan project (D141100004914003). Also with the help of the information engineering department to complete this paper.

References

1. Zhao, Z.-F., Wei, W.-F., Qiang, M.A.: A real-time processing system for massive sensing data. *Microelectron. Comput.* **29**(9), 10–14 (2012)

2. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. In: Proceedings of the SOSP, pp. 20–43 (2003)
3. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large cluster. In: Proceedings of the OSDI 2004, pp. 137–150 (2004)
4. Apache Hive. <http://hadoop.apache.org/hive/>
5. Pike, R., Dorward, S., Griesemer, R., Quinlan, S.: Interpreting the data: parallel analysis with sawzall. *Sci. Program. J.* **13**(4), 227–298 (2005)
6. Thusoo, A., Sarma, J.S., Jain, N., Shao, Z., et al.: Hive-a petabyte scale data warehouse using Hadoop data engineering. In: Proceedings of the ICDE, pp. 996–1005 (2010)
7. Olston, C., Reed, B., Srivastava, U., Kumar, R., Tomkins, A.: Pig Latin: a not-so-foreign language for data processing. In: Proceedings of the SIGMOD, pp. 1099–1110 (2008)
8. Chen, T., Xiao, N., Liu, F., Fu, C.S.: Clustering-based and consistent hashing-aware data placement algorithm. *J. Softw.* **21**(12), 3175–3185 (2010)
9. Patten, S.: The S3 Sookbook: Get Cooking with Amazon’s Simple Storage Service. Sopobo (2009)
10. Murty, J.: Programming Amazon Web Service: S3, EC2, SQS, FPS, and SimpleDB. O’Reilly, Sebastopol (2008)