http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (22): 6636-6640, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6636.6640

Organization and Management of Meteorological Sensor Network Collected Big Data

[1]Dong Wen, [2]Gu Guo-min, [1]Wu Tian-jun and [1]Yu Xin-ju
[1]State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. Beijing 100101, China
[2]Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China

Abstract: Global-scale meteorological observations and scientific research development has brought meteorological big data management and analysis bottlenecks. Aiming at this problem, this study focuses on the organization and management of meteorological sensor network collected big data in cloud computing environment based on the analysis of big data characteristics, application technology status and the main features and application requirements of meteorological sensor network data. And a meteorological sensor network big data organization and management framework based on the "storage-Calculation" integrated cluster environment is proposed. Particularly and focuses on the meteorological sensor network collected big data model in the cluster environment is put forward. The above framework and data model are put into practice in the meteorological data management and service system of Zhejiang province.

Key words: Big data management, meteorological sensor network, cloud GIS, cluster computing

INTRODUCTION

With the increasing prominence of the climate change issues, the scope, depth and accuracy of meteorological observation and analysis are proposed new requirements, which promote the rapid development of meteorological observations and analysis techniques. The development of modern sensor network technology has brought a dramatic increase in meteorological data, but the traditional GIS technology is difficult to achieve real-time access to this kind of spatial information, fast processing and comprehensive analysis and can not provide real-time, accurate and relevant spatial information services for all levels of users.

In recent years, following the cloud computing, Internet of things, big data technology is becoming another disruptive IT technology revolution. In the field of computer technology, the issues of big data definition, data characteristics and key technologies have become a hot research spot. But, it needs to develop the methods and technology system of massive spatial data for solving computational problems of massive meteorological sensor network data and similar information. The organization and management of big spatial data is an important part of the whole system.

STATUS OF BIG DATA MANAGEMENT

In the background of "Internet of Things ", "mobile Internet" and "cloud computing", the explosion of information gradually creates a "big data" concept. May 2012, UNGP (2012) project "Global Pulse" release "Big Data for Development: Challenges and Opportunities" to promote digital data and innovation of fast data collection and analysis methods. EOP, 2012b, In the "Big Data Research and Development Initiative" of the United States , the development and application of big data rose from the previous business practices up to the national strategic deployment, which greatly give impetus to the development technologies and tools of the Big Data (EOP, 2012a).

Research results at home and abroad mainly take advantage of distributed parallel computing to complete the application of traditional data processing techniques to large-scale distributed computing environments. These parallel computing modes including library computing, memory computing, data flow computing, distributed programming model are used to achieve big data computing for further improving the efficiency of data processing and analysis (Xi *et al*., 2009). For example, MapReduce is used to contract semantic data to reduce storage costs and improve data access efficiency

Corresponding Author: Gu Guo-min, Zhejiang University of Technology, Hangzhou, 310023, China

(Urbani *et al*., 2010). Scarlett (Ananthanarayanan *et al*., 2011) and BlobSeer (Nicolae *et al*., 2010) to optimized the data storage mode of MapReduce framework, scattered hot spots as much as possible to avoid excessive concentration of hot data storage which would affect the efficiency of data access. Hadoop (Bu *et al*., 2010) provides users with a simple interface which is easy to resolve complex jobs and run multi Job iterations.

Currently, big data related technologies can only be seen as a technology demonstration and it is difficult to address data management and the efficiency of data processing in different enterprises. It is necessary to establish their own technology systems for big data management in different areas.

CHARACTERISTICS OF METEOROLOGICAL SENSOR NETWORK BIG DATA

Meteorological sensor network big data has the basic characteristics of big data, what namely volume, variety, velocity, value and complexity. In addition, it also has the following special features.

First, the integration of multiple time scales. Meteorological sensor network includes monitoring station, meteorological radar, satellite and other sensing monitors and these sensing monitors could be used to monitor the same object from different spatial and temporal levels. So, monitoring data have different temporal characteristics. In addition, with the development of computer technology and professional analysis method, a variety of data products in different time scales have been produced on the base of original observation data. Because of these, users often need to integrate data which have different time scales in data application.

Second, continuous observation for a long time. Meteorological sensor network data come from continuous observation and simulation of meteorological elements or phenomena in reality, therefore, this data has a natural feature. At the same time, in order to meet the needs of study and application of meteorological science, we must meet the observation requirements in time length and frequency.

Third, "effective life" of data is short. Traditional data information is static and statistics information and the "effective life" time is also long. For example, basic terrain data often will not change much in a few years or even longer. However, meteorological sensor network data is accurate records of things' state at the time and represents the instantaneous state of things in theory. With the development of observation technology, the observation frequency of sensor is higher and higher and the "effective life" of meteorological sensor data of is shorter and shorter.

METEOROLOGICAL SENSOR NETWORK DATA ORGANIZATION AND MANAGEMENT

Data organization framework: In order to meet the demand for real-time data services, data organization and management needs to focus on resolving the problem how to complete the process from data access to data service in real time.

From the bottleneck of meteorological data management and application existing can be seen, the key to solve this problem is to realize mass meteorological data real-time processing and importing. However, on the conditions of the existing network and computer hardware and software technology, single computer or master-slave cluster have been unable to completely meet the needs of existing massive meteorological sensor data management. In order to maximize network and computing resources rate, this study uses the "storage-Calculation" integrated cluster environment as the basic environment and then build the organization and management framework of meteorological sensor network data using the flexibility of multiple nodes. The framework is shown in Fig. 1.

In the whole framework structure, data storage model is the core and data storage mode will directly affect the basic operating efficiency of data. The basic

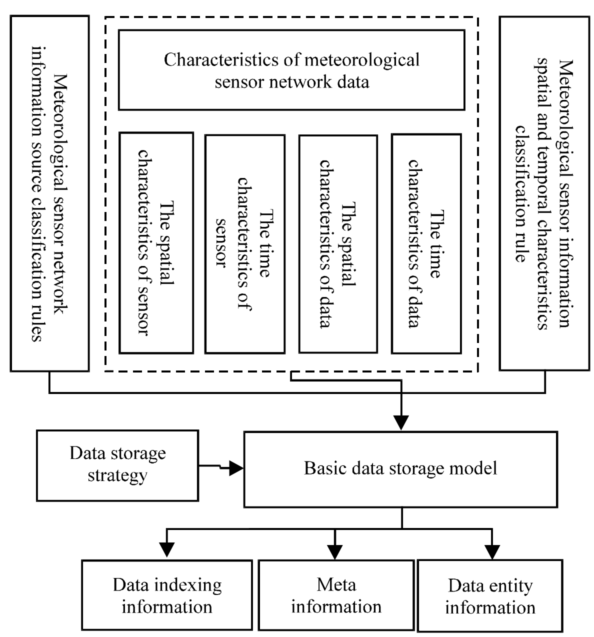

Fig. 1: Framework of organization and management of meteorological sensor network collected big data

Table 1: Basic sensor data model classification

No.	Type ID	Characteristics of sensor position	Spatial characteristics of data	Typical example
4001	DP	Definite position (D)	Point (P)	Observation station
4002	DS	Definite position (D)	Surface (S)	Radar
4003	MP	Mobile position (M)	Point (P)	Special monitoring
4004	MS	Mobile position (M)	Surface (S)	Satellite

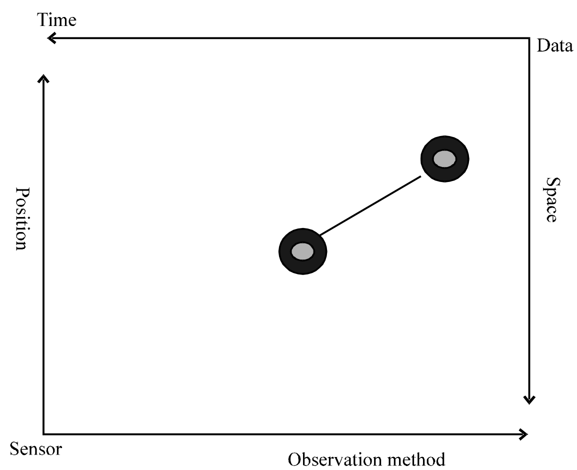

Fig. 2: Conceptual data model

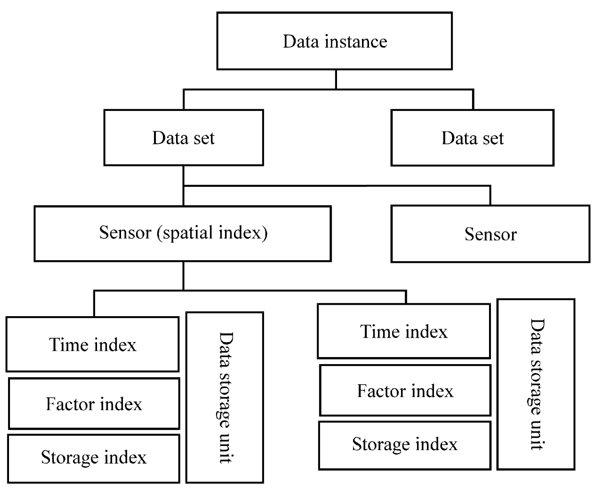

Fig. 3: Logic data model

meteorological sensor network data storage model is a set of data model which have been abstracted to represent the typical meteorological sensor data types. This classification is mainly based on the meteorological sensor network information source classification rules and meteorological sensor information spatial and temporal characteristics classification rules. Different types of data are stored in the corresponding basic data model. Each basic data model contains meta information, data entity information and data indexing information.

Data model: Data model is the core of meteorological sensor data organization and management framework. This study mainly focuses on the concept data model and logic data model.

Conceptual data model: sensor type is one of the most important factors of meteorological sensor data, which often determines the main features of observation elements. Therefore, meteorological sensor data could be described in two aspects- data expression and sensor expression.

As Fig. 2 shown, sensor and its observation data could be described from time and space two characteristics. Spatial characteristic of the sensor is mainly reflected in the position information of sensor and time characteristic is mainly embodied in the observation methods. Spatial characteristic of data is mainly reflected in the spatial range of data and time feature is mainly reflected in effective time range of data. Therefore, a type of meteorological sensor data could be described by a set of points in the following four dimension space.

Because of the time characteristic of data determined by the sensor observations, which was observation frequency and most meteorological sensors using interval observation methods, so, the basic meteorological sensor data model can be divided into the following categories, as Table 1 shown.

Logic data model: The logical data model is mainly expressed abstract levels of meteorological sensor data and the relationship between these levels, as Fig. 3 shown.

The data instance was used to achieve data organization and indexing at application and service level and different data sets were created under a data instance for the management and indexing of basic data classification. The data set contained data set metadata and sensor data and sensor data mainly included spatial and temporal characteristics about sensor, which was used to establish first level spatio-temporal index. Each kind of sensor data has different organization, storage and management method, but is mainly composed of second level data index and data storage unit and the second level data index mainly contains temporal characteristic index, data element category index and storage location index. This study designs the data storage unit structure and index methods for different

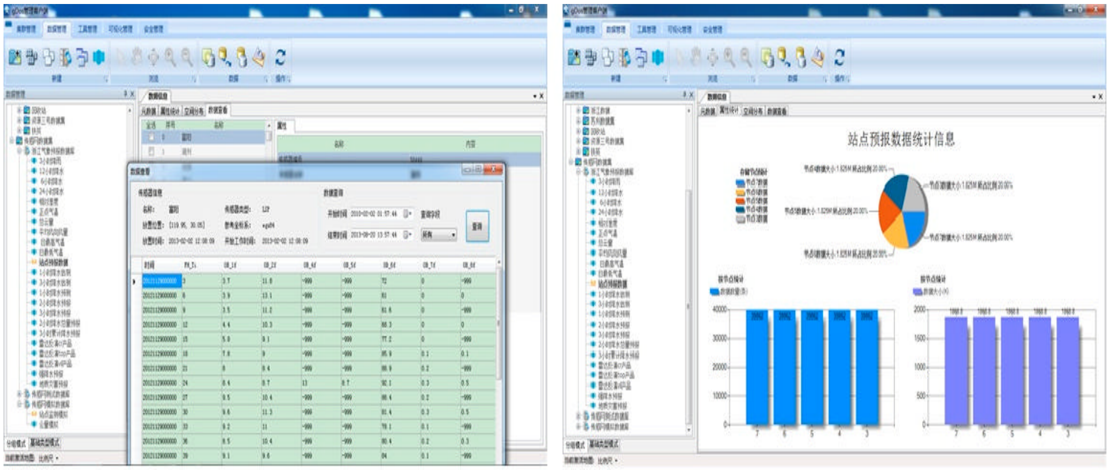

Fig. 4: Metadata management and monitoring site distribution management

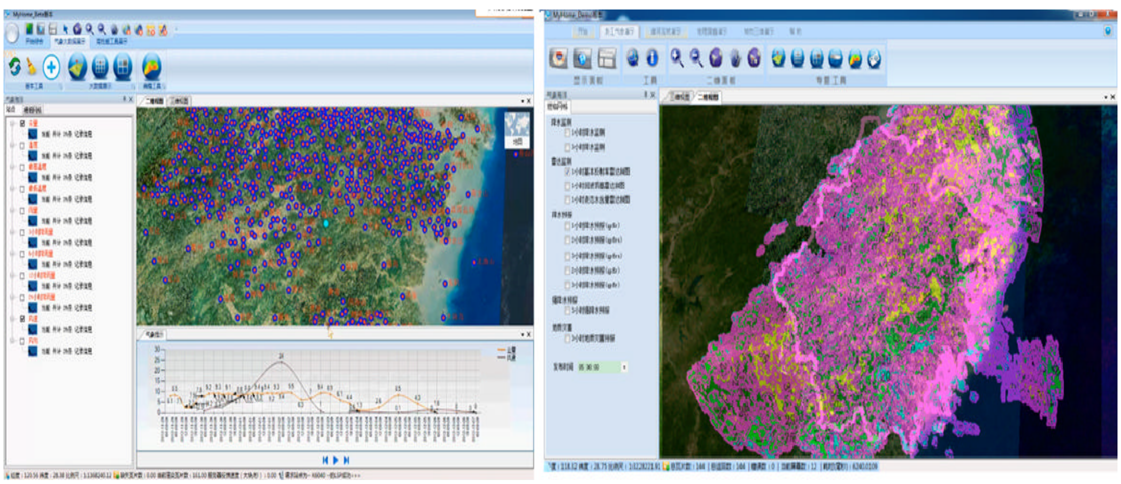

Fig. 5: Sensor distribution and contour analysis result

types of data and constructs the meteorological sensor data model on the basis of basic classification.

Different abstract levels of the logical data model would be implemented in the cluster server and cluster nodes respectively in the "storage-Calculation" integrated cluster environment, considering the data storage, computational and network efficiency and other factors. The cluster server completes the organization and management of data instances, data sets, sensor information and the first level index of data and nodes mainly realizes the storage and management of data and the second level index of data.

APPLICATION

The meteorological data management and service system of Zhejiang province was built on the basis of the cloud GIS platform and this system provide real-time, efficient management of all kinds of meteorological sensor network data, online data service and data analysis function for different users.

The meteorological sensor network data organizing framework and data model were put into practise in this system and some professional data service and related data analysis functions were provided based on this, such as meta information management, data distribution management, sensor distribution, sensor data analysis and so on. Figure 4 show the meta information management and data distribution management results and Fig. 5 show sensor distribution and sensor data analysis functions.

CONCLUSION

This study puts forward a data organization and management framework of meteorological sensor network big data in the "Storage-Calculation" integrated cluster environment and focuses on the data model. Finally, this framework and data model were implemented in the meteorological data management and service system of Zhejiang province. This system can satisfy the basic meteorological data management requirements, quasi real time data access and service and real-time online analysis of 1 minutes frequency observation data, which means this framework and data model can be used to solve the big data management problem effectively and provides good data service for real-time meteorological analysis.

ACKNOWLEDGMENT

The authors would like to thank for the support by Natural Science Foundation of China under the Grant 41301438 and the support by China Academy of Sciences major projects under the Grant KZZB-EW-07-02 .

REFERENCES

Ananthanarayanan, G., S. Agarwal, S. Kandula, A. Greenberg, I. Stoica, D. Harlan and E. Harris, 2011. Coping with skewed content popular-ity in MapReduce clusters. Proceedings of the 6th European Conference on Computer Systems, April 10-13, 2011, Salzburg, Austria, pp: 287-300.

Bu, Y., B. Howe, M. Balazinska and M.D. Ernst, 2010. Haloop: Efficient iterative data processing on large clusters. Proceedings of the VLDB Endowment, Volume 3, September, 2010, Seattle, WA., USA., pp: 285-296.

EOP, 2012. Fact sheet: Big data across the federal government. March 29, 2012, Executive Office of the President, USA.

EOP, 2012. Obama administration unveils big data initiative: Announces $200 Million in new RandD investments. March 29, 2012, Office of Science and Technology Policy, Executive Office of the President, USA.

Nicolae, B., D. Moise, G. Antoniu, L. Bouge and M. Dorier, 2010. BlobSeer: Bringing high throughput under heavy concurrency to hadoop map-reduce applications. Proceedings of the 24th IEEE International Parallel and Distributed Processing Symposium, April 19-23, 2010, Atlanta, GA., pp: 1-11.

UNGP, 2012. Big data for development: Challenges and opportunities. May, 2012, United Nations Global Pulse, USA., pp: 13-20.

Urbani, J., J. Maassen and H. Bal, 2010. Massive semantic web data compression with MapReduce. Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, June 21-25, 2010, Chicago, IL., USA., pp: 795-802.

Xi, J.Q., J.G. You, D.Y. Tang and W.J. Xiao, 2009. A parallel closed-cubing algorithm based on MapReduce. J. South China Univ. Technol. (Nat. Sci. Edn.), 37: 91-112.