Journal of Electronics and Information Science (2020) 5: 63-71
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2020.51010
ISSN 2371-9524

Construction of distributed Data management platform for Land Engineering based on Hadoop Framework

Hui Kong[1, 2, 3, 4]

[1]*Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China*

[2]*Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710021, China*

[3]*Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710021, China;*

[4]*Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710021, China*

Keywords: Land engineering, Big data, Hadoop framework, Distributed

Abstract: This article to build land engineering needs of the business platform for the large data as an opportunity, from a computer application system design method and the principle of the premise, from the overall system architecture, logic structure design, database management system design, the upper application function design and data visualization research and development, using the Hadoop distributed system architecture technology, Web application technology and GeoServer space information technology, based on "a structured and unstructured + space" mode of distributed data organization and management, Based on the application system implementation mode of "Web+GIS",combined with the classic MVC development mode and B/S (browser/server) multi-layer architecture mode design, the land engineering big data platform integrates five modules, including thematic data management, data retrieval, simulation, decision analysis and system comprehensive management.

1. Introduction

At present, the social development speed, is in a rapid comprehensive integration, data mining, and application as the main characteristics of the intelligent phase depth fusion, in the man - machine- the backdrop of the ternary blend, to "all things must be connected, anything can be programming" as the goal, digital, network and intelligent integration development new situation. In the face of the diversity of land engineering business needs and the particularity of land engineering science research, it is also urgent to realize the effective management of land engineering industry application and discipline research through land science and technology innovation and seeking for more advanced technical means. The construction and application of land engineering big data platform is an important direction of land and resources management in the era of intelligence. The development of the land engineering industry is a measure to implement the national comprehensive land improvement strategy. However, the existing technological means cannot meet

the development needs of the land engineering industry. The introduction of key big data technologies into the research and application of the land engineering discipline has huge development potential.

Therefore, this article to the business requirements in land engineering data organization management is weak, and low level of data encapsulation, by integrating Spring business logic layer and data layer of loose coupling, Tomcat lightweight Web applications and AJAX application layer and business logic layer interaction middleware technology in the platform layer for heterogeneous data encapsulation, integration, exchange and sharing, innovation in service composition pattern, improve the land of the dynamic characteristics of large data application and scalability. By combining with B/S architecture, such as WMS, WFS, data standard, the service side GeoServer space technology, unified the various types of user interaction and interface standards in the application layer, solves the distributed heterogeneous data, connectivity and efficient use of technical means, to build the application of rapid prototyping land engineering data platform, has realized the land utilization of engineering data.

2. Overall platform design

To land engineering key technology and application research as a starting point, large data structures, land engineering big data platform computer room environment and hardware system as the foundation, through the data collection, data storage and management, and analyzes the restructuring of the data to build land engineering database, and based on SOA architecture system and network service platform structures, an application in the management platform, the system of the whole design idea is shown in figure 1.

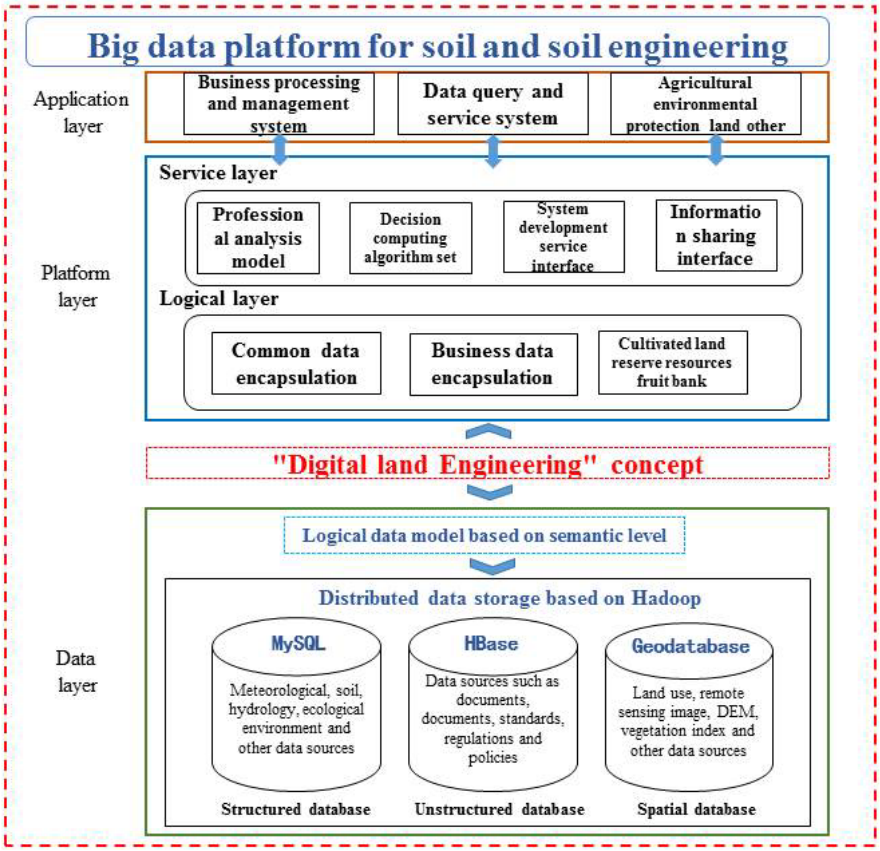

Figure. 1 Overall system design

This system involved in land engineering industry starting point, on the basis of the total factor data based on JavaScript technology, the Java programming language, for SpringMVC framework of the Web server, online map API, using B/S multilayer architecture and WebGIS technology for system design, to build a project data management, data retrieval, simulation, decision analysis, and comprehensive management system with functions such as land engineering data platform, provides users with data storage, processing, management and application, visualization, and other functions, Meanwhile, it also provides a convenient and efficient platform for relevant business personnel to effectively manage data. The data platform based on open service-oriented Architecture (SOA) is a multi-functional application system based on Web technology, GIS space technology and B/S architecture. It consists of three layers of data layer, logic layer (business layer) and application layer. It is mainly based on the data layer and core, relying on the network platform and various services to support the basic, through the management and presentation of data to efficiently monitor data and analysis data.

3. System database design

The database design idea of land Engineering big data platform is based on the real-time business data and is implemented based on the typical three-tier database architecture pattern, namely "physical layer", "logical layer" and "expression layer". Physical layer is the concept layer in the database design, including the composition of the overall business database, because the data based on the system is relational data, so it has also been based on the relational database, to achieve the construction of structured database, unstructured database, spatial database. The logical layer is mainly aimed at the design of the main table, schedule and connection relationship of specific data table. Based on the database design paradigm, the logical relationship between each table is constructed, and the data structure and attribute fields of each table are composed. The expression layer is through the database add (Insert), Delete (Delete), Update (Select), data front and background interaction, and in a certain form in the front-end expression.

3.1 Composition of database

In the big data platform of land engineering, we design the system database by using the basic database of data collection and storage, data processing and analysis, data application in display equipment.

(1) Structured database

Structured database is mainly of special libraries currently include meteorological projects, soil, river projects, air quality, legal projects, land use, water projects, mineral projects, groundwater, ray radiation, forestry projects, such as 11 parts, each project is on the related data collection and display itself. The data table is designed for the parameter information and index of the acquisition equipment and the surrounding environmental monitoring equipment, including the standard two-dimensional structure of meteorology, hydrology, soil, ecological environment, etc. The MySQL relational database is mainly used as the main technical means to achieve this.

(2) Unstructured database

Unstructured database is the database management of all non-two-dimensional table structures involving documents, documents, standards, regulations and policies. It is mainly realized by Hbase column database.

(3) Spatial database

Spatial database has vector and raster data management for some spatial element information, including image data of map service, some files, diagrams, tables and other data management, which is realized by Geodatabase spatial data management mode.

3.2 Conceptual design of database

Database conceptual design refers to the process of abstracts the user requirements obtained from the previous requirements analysis into information structure, that is, conceptual model. It is the key to the entire database design, as shown in Figure 2.

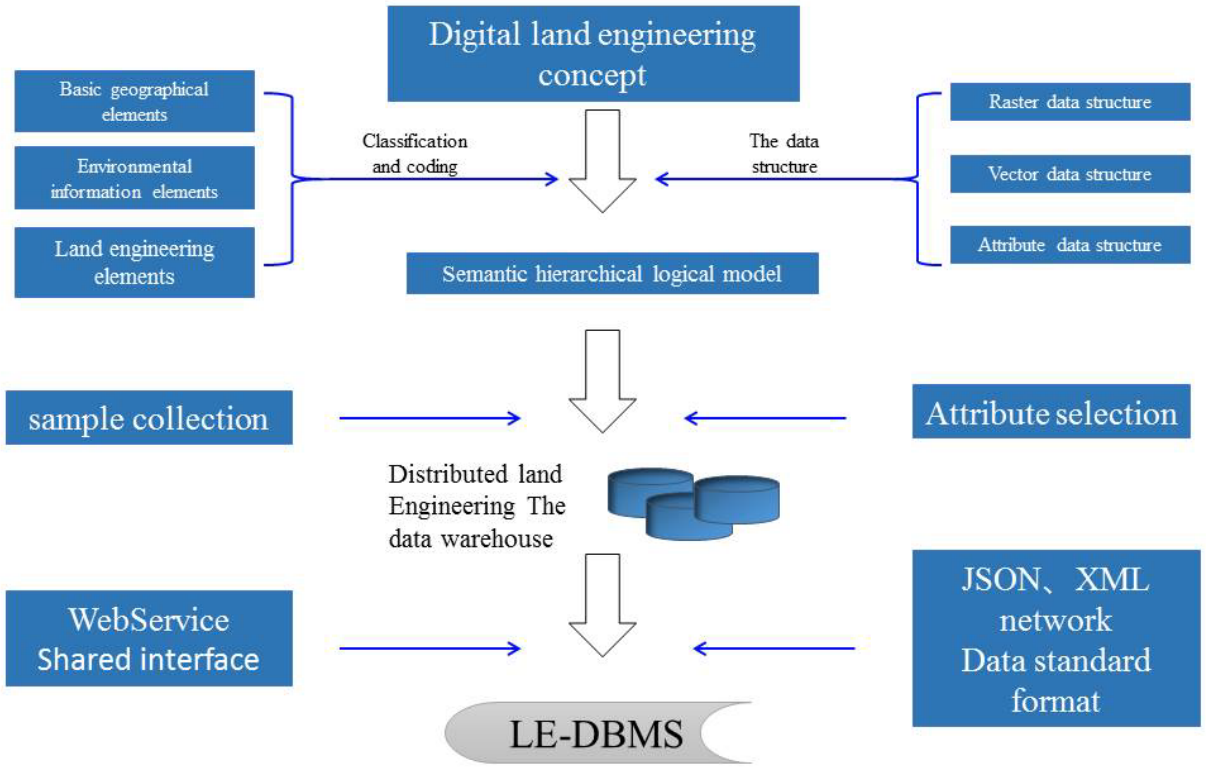

Figure. 2 The database concept design

3.3 Detailed database design

Detailed database design is to design each table structure of the basic database of land engineering big data platform, including table name, table field name, and field type, primary key, foreign key, constraint and other database standards. There are many database tables designed on this platform. For the sake of the expression of this article, it is not enough to enumerate one by one. This time, meteorological data source is taken as an example, as shown in Table 1 and Table 2 below.

Table 1 Monitoring site database tables

Data table ID number	Table name	note
1	ZTSJ_QX	Meteorological data source
2	ZTSJ_SW	Hydrologic data source
3	ZTSJ_TR	Soil data source
…	…	…

Table 2 Monitoring equipment detailed design table

Field identification	Field names	null value	primary key	field type	precision
ZTSJ_QX_XH	The serial number	No	No	Int	10
ZTSJ_QX_JCZMC	Monitoring station site	No	No	Char	14
ZTSJ_QX_JCZDD	Name of monitoring station	No	No	Char	14
ZTSJ_QX_JD	longitude	No	No	Char	14
ZTSJ_QX_WD	latitude	No	No	Char	14
ZTSJ_QX_SBXH	precipitation	No	No	Char	20
ZTSJ_QX_CSPC	The temperature	No	No	Char	12
ZTSJ_QX_DY	The direction of the wind	No	No	Num	8
ZTSJ_QX_GNYT	The wind speed	No	No	Char	20
ZTSJ_QX_HB	At an altitude of	No	No	Num	8

4. System function module design

System function module is an important part of the whole land engineering big data platform operation, this platform analysis and design according to the requirements of the early stage of the target, and the corresponding function requirements, system design for five major functional modules, respectively is project data management, data retrieval, simulation, decision analysis, and comprehensive management system. According to the analysis and design of the data table structure in the previous chapter, the function module of this system is divided into two modules: data management module and application module. Data management module is a unified management of all kinds of data sources collected and added to achieve the management of the entire system, including data query, data management and comprehensive management (including user information, password modification, authority management, etc.).Application module is the sum of all modules providing services for the entire user, including simulation (soil-crop simulation process management and simulation), data retrieval (retrieval of various thematic data and related data), and decision analysis (including analysis and statistics of solid data based on relevant model algorithms).

4.1 Platform layer function design

It is mainly a big data distributed data storage and analysis platform built on the basis of Xdata-SDH system and Xdata-AUS system based on Hadoop standard system. The overall platform design architecture meets the following features: high reliability; adopt cluster technology, multi-copy technology, independent backup technology and other means to effectively reduce the data failure rate, and build a reliable data application service platform;

(1) High performance: the distributed cluster technology is adopted to improve the retrieval capability of the core database, so as to meet the user's requirement of parallel processing without delay or downtime during peak business hours;

(2) Forward-looking: the system planning and design have a certain leading and forward-looking awareness, to ensure that the distributed core database in a relatively long period of advanced and stable, and at the same time have a good upgrade ability, so that the distributed core database can smoothly achieve a smooth upgrade to the newer generation of equipment, technology platform;

(3) Scalability: Using distributed storage technology to realize dynamic expansion of data nodes and automatic load balancing; Complete the node capacity expansion without affecting the system work.

(4) Maintainability: A set of xdata-SDH and Xdata-AUS interactive operation and maintenance management software is provided, to form a unified management mode of distributed system. Managers can monitor metrics and status information such as overall storage services, storage capacity, and operate the cluster in all directions. Through big data platform, data fusion and analysis and calculation are completed for users, which lays a solid foundation for users to carry out data informatization construction.

4.2 Function design of data layer

The data layer involves a complete process of data collection and summary for the relevant business data required by the user.This layer includes data acquisition system, data processing system and data storage system, which respectively realize data acquisition, collection and transformation, and data storage, as shown in Figure 3.

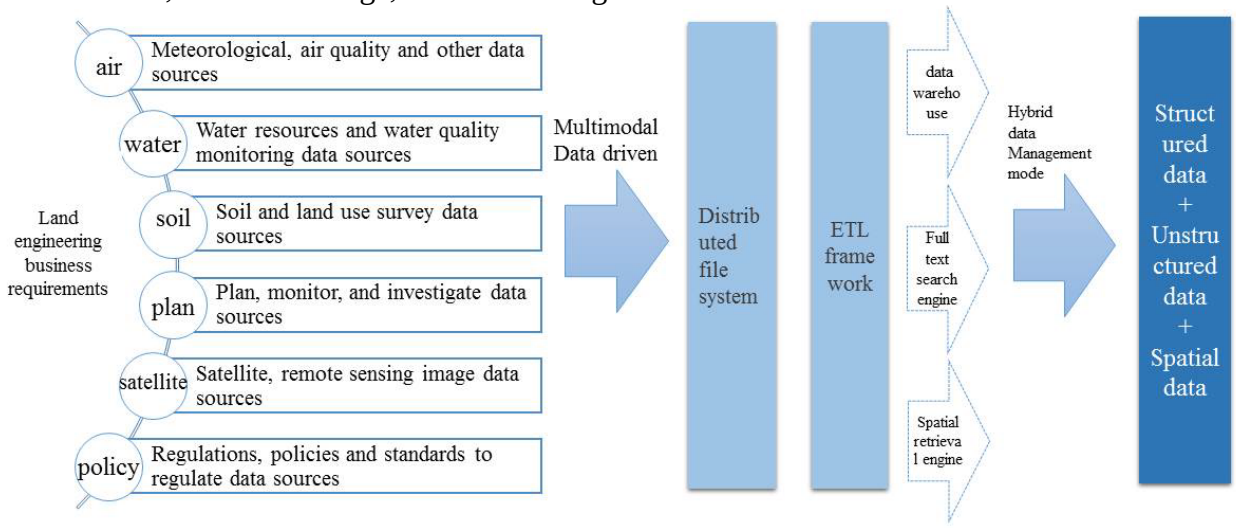

Figure. 3 Data Layer Functional Design Drawing

(1) Data acquisition system

Platform for business related to weather, soil, land use, legal, water, air quality, a variety of categories of data, such as the solar radiation across a number of areas and departments, which have already bought the data through the database interface or text processing program to complete data access, but these data do not fully meet the needs of the business. Therefore, for the incomplete data, it is necessary to extract and collect the information published on the website of relevant departments in combination with web crawler technology.

(2) Data processing system

On the basis of collecting all kinds of data from the internal and the network, the data classification and processing should be carried out in the later stage for each data content, so as to meet the requirements of the application layer. It is mainly oriented to different formats such as structured data, text data, spatial data, etc., and it converts the data according to the purpose and usage scenarios of each type of data. After the data is converted into a unified standard format, special cleaning operations are required for the outliers or error messages of each type of data, and the outliers are supplemented by difference or average values to support the implementation of the upper logical analysis function.

(3) Data storage system

Data storage is mainly based on the characteristics of all kinds of data on each data platform to

establish a unified database table structure, and through metadata management to ensure the speed and efficiency of data extraction strategy. Currently, structured statistics and text content are stored in distributed nodes, while text data is stored in the HDFS file system. To ensure data security, each distributed node and HDFS system ensures data security through replication mechanism, and defines a snapshot of the completed data and backs it up to HDFS. The stability and security of data guarantee the extraction and analysis effect of data service layer and data application layer.

4.3 Function design of service layer

The service level provides the analysis interface and computing services for the caller's multiple data to the platform data and computing resources.

(1) Spatial data: Publishing WMS service in GeoServer enables users to extract and invoke spatial data through HTTP protocol according to specific business requirements.

(2) Structured data: While providing Java standard invocation interface and Restful service interface, it provides users with standard SQL query syntax query support, and provides analysis engineers and data callers with cross-language and cross-node data analysis services.

(3) Algorithm service: Unified scheduling and management of platform resources can be realized through YARN (Yet Another Resource Negotiato);Publish algorithm services with MapReduce and Spark computing frameworks, integrate Mahout, MLib and other algorithms libraries, and provide continuous and efficient services for upper-layer algorithm analysis.

4.4 Function design of application layer

The upper application of the platform mainly includes data retrieval, simulation and decision analysis.

(1) Thematic search

Thematic retrieval mainly forms a corresponding visual thematic display of data on the internal data of the platform to facilitate data query and retrieval by users.

(2) Simulation

Simulation provides cloud computing services based on mature mechanism models, provides users with online simulation of the whole set of crops, and saves corresponding simulation data to provide data support for later simulation research.

(3) Decision analysis

The decision analysis is mainly aimed at the current business, which integrates the data of multi-industry platform in time and space, and establishes a series of regional planning and decision means, such as regional planning, input of decision information, urban decision making, comprehensive analysis, crop recommendation, soil reconstruction, etc.Combined with big data technology, reliable and stable data theoretical support is provided for the business implementation process of users, as shown in Figure 4.

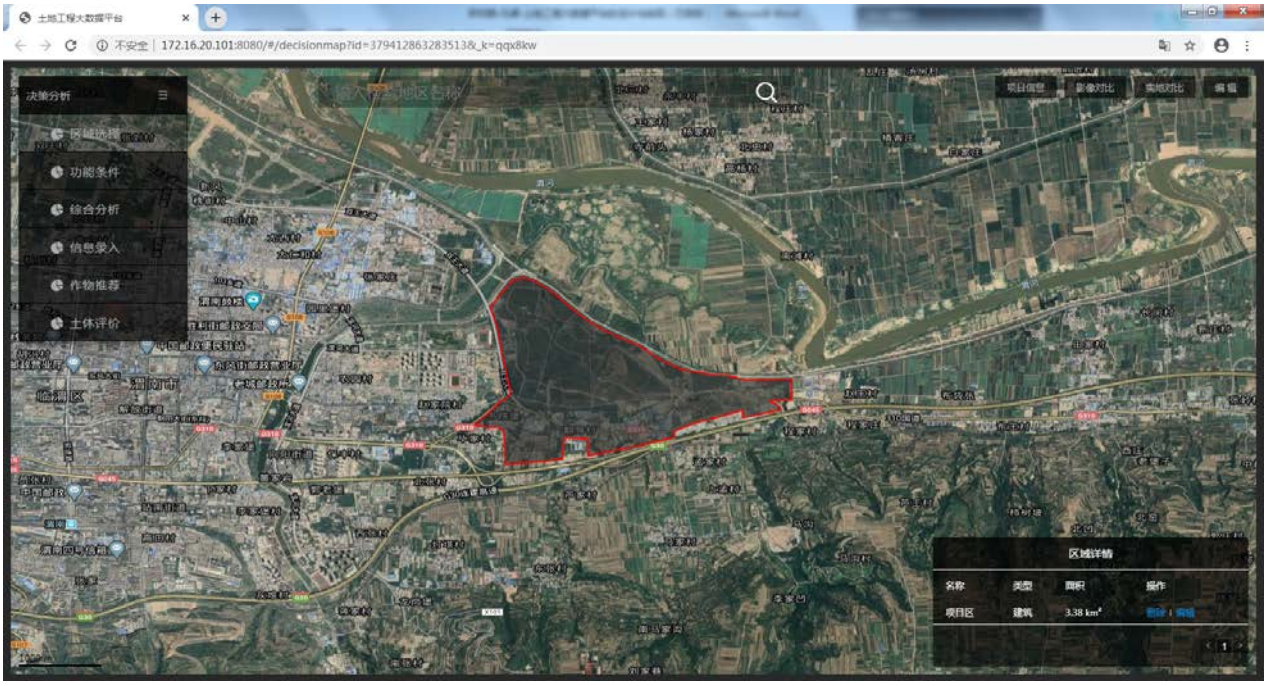

Figure. 4 Decision analysis application system interface

5. Conclusion

Through the extensive theory research, technology practice, through the combination of big data technology and land project, the logical data model using semantic levels, constituting land engineering industry is suitable for classification and coding of data elements and data structures, by combining the Hadoop distributed database, MySQL relational database, database Geodatabase space type, set up many modal data management, data in a structured and unstructured data and spatial data sharing classification model for digital interface, distributed storage system is adopted to establish the data warehouse, Land engineering data management system (LE-DBMS) is constructed. The first land engineering big data platform was built by integrating the loose coupling technology of Spring business logic layer and data layer, the lightweight Web application technology of Tomcat and the interaction middleware technology of AJAX application layer and business logic layer, and the new generation of information technology was successfully applied to land engineering practice. The database of land engineering is established and a new application mode of land engineering is developed. The work to build the first land engineering database, the land and natural resources management informationization means ascension has important practical significance, to the construction of database, the data acquisition system to improve the practical application, the project has larger breakthrough, but in different kinds of renovation project subsystem of research and development and the extension application has yet to be promoted, it will be the results of the study in-depth development goals and direction in the future.

Acknowledgements

Thanks for Research on Quick Calculation of Earthwork in Land Consolidation Based on Digital Elevation Model Data (2018-TD02) and Research on Quick Calculation of Earthwork in Land Consolidation Based on Digital Elevation Model Data (DJNY2019-29).

References

[1] Zhang Feng. Seizing the new opportunity of big data development to promote the healthy development of big data [J].China informatization, 2016 (7): 9-11.
[2] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey. Business intelligence and analytics: From Big Data to Big Impact [M]. Society for Information Management and Information Systems Research Center, 2012.
[3] Han Jichang.Introduction to Land Engineering [M]. Beijing: Science Press, 2013.
[4] Jin Xianfeng.Zhang Zelie. Thinking on informatization Construction of Big Data Era Planning [J]. Planner, 2015 (3): 135-139.
[5] CAI Lihui.Constructing digital land and promoting the informationization process of land resource management [J]. Journal of sun yat-sen university (social science edition), 2003, 43 (6): 62-69.
[6] Gong Xiayi, Li Bohu, Chai Xudong, et al. Overview of Big Data Platform Technology [J]. Journal of System Simulation, 2014, 26 (3): 489-496.
[7] Ji Qianqian. Research on service Mode of Public Transport Big Data Platform Architecture [D].
[8] Fei Shiyi. Collaborative Research between Hadoop Big Data Platform and Traditional data Warehouse [D]. 2014.
[9] Naimi A I, Westreich D J. Big Data: A Revolution That Will Transform How We Live, Work, and Think [J]. American Journal of Epidemiology, 2014, 179 (9): 1143-1144.
[10] Yue Wenze.Land Big Data Development and Land Management Transformation [J]. Land Observation, 2015.