Applied Research on Agricultural Big Data

Feng Wang
Party School of Liaoning Provincial Party Committee, Shenyang 110004, Liaoning, China

Email: wfengon@163.com

Abstract. With the continuous advancement of digital informationization, the agricultural field has naturally entered the era of big data. At present, big data technology has achieved remarkable results in the Internet industry, but it is still in its infancy in the agricultural field. Due to the complexity of the agricultural field itself and the obvious diversity and heterogeneity of data, the agricultural Internet of Things in the big data environment is facing many difficulties and challenges, especially the collection and storage of massive data. The purpose of this article is to study the application of agricultural big data. This article briefly introduces the background of the development of agricultural big data and the significance of developing agricultural big data, and summarizes the meaning of big data. At the same time, this article summarizes and analyzes the major applications of agricultural big data in production process management, agricultural resource management, and agricultural ecology environmental management, agricultural products and food safety management. In the experimental part, this paper designed and implemented an agricultural big data system platform based on Hadoop. This article simulated the query of 10, 100 and 1000 records respectively for 10,000, 100,000, 1 million and 10 million records, and the average response time was 56ms. The experimental results show that the performance of the agricultural big data platform designed in this paper can meet the needs.

Keywords: Agricultural Big Data, Big Data, Agricultural Informatization, Hadoop Platform, Performance Testing

1. Introduction
The agricultural big data platform plays an important role as an important carrier of agricultural information resources. Sorting out and developing a platform with functions for managing and analyzing agricultural big data is the key to solving the current inefficient use of big data in the agricultural field. The Chinese Academy of Agricultural Sciences put forward the development of agricultural informatization, and the development of agricultural big data requires research on agricultural data analysis technology to find a reasonable and efficient analysis method. If you want to manage, integrate, and analyze the large amount of data generated by agricultural informatization, you need to develop an application platform with higher technology. These technologies need to support the platform to realize data collection, data storage, data processing, data analysis, data display and
other functions. Only with the support of these technologies can the agricultural big data platform better play the role of data; only the platform deeply couples big data technology with the agricultural field, can we monitor and alert China's agricultural market, and can achieve smart agriculture to guide production management, can make scientific decisions on the direction of China's agricultural development: Only under the support of key technologies, can agricultural data be collected, standardized, publicly released, and forecasted to affect expectations and manage future agriculture. Therefore, the key technology for the development of agricultural big data platforms and research platforms is the key points and hotspots of rational use of agricultural big data.

With the development of agricultural informatization, many agricultural-related websites, higher agricultural institutions, and agricultural-related government departments have collected and stored a large amount of agricultural data [1-2]. Including various agricultural data updated daily by the agricultural website, experimental research data used by scientific research in universities, and agricultural census statistical data [3]. These massive agricultural data are waiting to be developed and used, but most of them are buried in the database and data room, and the potential value of the data is lost [4]. If a fully functional platform is developed, corresponding key technologies are applied, and agricultural data is managed, analyzed, and used in a unified manner, it will provide data assistance for agricultural development and provide guidance and recommendations for agricultural production [5-6]. Such a platform can not only integrate these data, but also the management function of the platform can complete the data exchange function, which enriches the data volume of the platform [7]. Become a system that integrates massive agricultural data, and then process and analyze the key technologies of the platform to provide the most comprehensive scientific research data for governments, schools, etc., provide farmers with agricultural production knowledge, provide the government with correct agricultural development decisions, and accelerate the process of China's agriculture to digital transformation [8-9].

This article briefly introduces the background of the development of agricultural big data and the significance of developing agricultural big data, and summarizes the meaning of big data. At the same time, this article summarizes and analyzes the major applications of agricultural big data in production process management, agricultural resource management, and agricultural ecology environmental management, agricultural products and food safety management. In the experimental part, this paper designed and implemented an agricultural big data system platform based on Hadoop.

2. Method

2.1 Big Data
Big data does not have a fixed standardized definition. Even so, it will not deny its effect on social development, but it has different values in different countries for different objects [10-11]. Its core role is irreplaceable. The diversity and differentiation of data determines that it can make accurate judgments according to different environments, which contains great application value. Most people are familiar with structured data, but what is really used in practice is unstructured data. For example, text processing and e-mail transmission are semi-structured data. It belongs to the transition period. The real big data is the Internet and e-commerce types, such as documents, web pages, reports, pictures, audio, and video[12].

2.2 Application of Agricultural Big Data
At present, in the development of agricultural information in China, big data plays a key role. Through the analysis of data sources, China currently applies big data in the following areas: Production process management data: The first step in the use of agricultural information lies in production. The link involves aquaculture, planting and agriculture. In the production process, we must first ensure that the data can accurately detect the multi-factor environmental factors before planting. Second, we must have a decisive and intelligent decision-making plan after production. Finally, we must make a reasonable allocation in regulation and management. Issues to be resolved as soon as possible;
agricultural resource management data: sustainable development is the most basic prerequisite for economic growth. Compared with the last century, China's agricultural resources are scarce, so we must continue to use the data in accordance with the current development situation and the diversity of the ecological environment to use data. Improve agricultural production, overall planning, and improve the effective reuse of agricultural resources; agricultural ecological environment management data: soil, atmosphere, water quality, meteorology, pollution, disasters, etc. Comprehensive monitoring and accurate management are needed; Big data on agricultural and food safety management: The ultimate purpose of agricultural production is to meet the most basic guarantee of survival for the people. The safety of products has become a major concern for consumers. In order to reduce costs, farmers spend all their efforts on agricultural products and undermine food safety. According to this situation, relevant agricultural departments should strengthen the management of the environment of crop production areas, production workshops and market circulation areas; big data on agricultural equipment and facility monitoring: agriculture the production environment should be matched with the resources and the market, and appropriate monitoring should be conducted to facilitate the processing of agricultural information.

2.3 Hadoop
Hadoop is a software platform for the development and operation of large-scale data. It is an Apache open source software framework implemented by the Java language. Users can develop distributed programs without knowing the underlying details of the distribution. Make full use of the power of the cluster for high-speed computing and storage. Hadoop implements a distributed file system (Hadoop Distributed File System), referred to as HDFS. HDFS has high fault tolerance and is designed to be deployed on inexpensive hardware; moreover, it provides high throughput to access application data and is suitable for applications with very large data sets. HDFS relaxes the requirements of POSIX and can access data in the file system in the form of streams. The core design of the Hadoop framework is: HDFS and MapReduce. HDFS provides storage for massive data, and MapReduce provides computation for massive data.

3. Experiment

3.1 system Architecture
As a big data system, the first problem that needs to be solved is the data capture problem, then the data is cleaned and preprocessed, then the data is analyzed offline with Hadoop, and finally the visualization software is used to display the results of the data analysis using Echarts. The system contains five main modules: data crawling module, data cleaning module, Hadoop offline analysis module, Hbase database persistence module, and Web data visualization module. At the lowest level is the data scraping module. Agricultural big data is mainly obtained from agricultural big data application cloud platforms, Blake agricultural data and other websites. After the original data is obtained, it needs to be sorted and cleaned to remove those unreasonable values; then, the offline analysis phase is entered. This phase is mainly based on the Hadoop MapReduce programming model, writing multiple MapReduce tasks, and storing the final results locally. Then the data persistence phase, that is, using the MySQL database, the text file after offline analysis is parsed, and then stored in the MySQL database; the web data visualization module is directly facing the user, and the user can easily see through this web interface to data analysis results. Data reports include the use of histograms, histograms, pie charts, to display data analysis results.

3.2 Experimental Environment
The experiments in this paper mainly build a Hadoop development and debugging environment on the Windows 10 operating system, and then use vmware14.0 software to build a distributed Hadoop environment.
4. Discussion

4.1 Experimental Results and Analysis
For the read performance test, when 10,000, 100,000, 1 million, and 10 million records have been written to the data table, test the query for 10, 100, and 1000 records from the data table. Required system response time. The experimental results are shown in Table 1 and Figure 1:

Table 1. Experimental results

| Data record (ten thousand) | Query the records | Response time (ms) |
|----------------------------|-------------------|--------------------|
| 1                          | 10                | 41                 |
| 1                          | 100               | 53                 |
| 1                          | 1000              | 62                 |
| 10                         | 10                | 51                 |
| 10                         | 100               | 50                 |
| 10                         | 1000              | 46                 |
| 100                        | 10                | 44                 |
| 100                        | 100               | 46                 |
| 100                        | 1000              | 86                 |
| 1000                       | 10                | 46                 |
| 1000                       | 100               | 54                 |
| 1000                       | 1000              | 93                 |

![Figure 1. Experimental results](image)

4.2 Suggestions for Promoting the Development of Traditional Agriculture
(1) Intelligent control
The traditional agricultural cultivation is based on the analysis and summary of the experience accumulated by the predecessors for a long time in production, and the conditions are more conducive to the growth and development of crops. This requires a long process, and due to the influence of uncertain factors, statistical data can sometimes be leaked. The use of big data can effectively provide scientific information for all aspects of smart agriculture. For example, when planting rice, first of all, research on the factors required for the growth of rice before, get the most scientific data, and choose the most suitable soil conditions according to the data. During the planting process, data analysis is performed in real time, and the growth situation is fed back to professionals. It can also be directly
used in accordance with the orders issued, such as irrigation and pesticide application. When the designated task is completed, it can be automatically stopped, and intelligent control and agricultural automation can be used to achieve more scientific planting.

(2) Prevention of agricultural information

Traditional agricultural production is affected by various factors such as market demand and natural conditions, and some unexpected situations will inevitably occur. Big data can effectively avoid agricultural risks by collecting information about agricultural production activities, market supply and demand, and real-time monitoring and appropriate response measures. At the same time, smart agriculture can also make predictions on the future development of agriculture, predicting some abnormal changes and preparing for the rain. Smart agriculture can avoid various problems by sorting out and analyzing the agricultural development situation in the past years and making effective judgments.

(3) Management aspects of agriculture

First of all, to achieve the goal of intelligently managing agricultural production. Although agricultural production is not fast-changing, people can be at the forefront of change. Smart agriculture must do a good job of collecting and sorting big data information in real time. It is necessary to conduct a specific analysis of a specific problem, provide a reasonable and effective agricultural formula, and use it in all aspects. Secondly, intelligently coordinate agricultural resources. Agricultural resources mainly refer to various resources related to agriculture, including land resources, water resources and climate resources. China's population base is large, and agricultural resources per capita are relatively small. In addition, agricultural resources are changing, and water, land and biological resources available for agriculture are becoming less and less. At this time, you can use big data to conduct overall surveys, integrate agricultural resources, and optimize the resource structure. In addition, you can use big data to manage agricultural resources in an integrated manner to ensure overall utilization and achieve high agricultural production and energy saving. Again, protect the ecological environment intelligently. Agro-ecological environment refers to the relationship between agricultural organisms and their relationship with the environment, and is an important part of the natural environment. Using big data, the overall measurement and dynamic supervision of all environmental elements can be achieved to achieve the effect of refined management. Finally, intelligentization guarantees the safety of agricultural facilities. Use big data to grasp the situation of facilities and equipment needed for agriculture in real time to ensure the safety of facilities.

5. Conclusion

The development of agricultural economy is not only closely related to people's daily life, but also affects the growth and development of the national economy. Therefore, in the process of agricultural development, we must realize the importance of big data, and innovate and optimize the application of big data in agricultural development, so as to ensure the sustainability of agricultural development. Big data technology has become a new driving force for the development of various industries in this era. Relying on big data technology to achieve technological change and innovation has become a must for all industries. The agricultural field also needs to rely on big data technology to realize the modernization of agriculture.

References

[1]. Alex H. Poole, Deborah A. Garwood. “Natural Allies”: Librarians, Archivists, and Big Data in International Digital Humanities Project Work[J]. Journal of Documentation, 2018, 74(4):804-826.

[2]. Lin Qiu, Sarah Hian May Chan, David Chan. Big data in social and psychological science: theoretical and methodological issues[J]. Journal of Computational Social Science, 2018, 1(1):59-66.

[3]. Smith SM, Nichols TE. Statistical Challenges in "Big Data" Human Neuroimaging[J]. Neuron, 2018, 97(2):263.
[4]. Xiaofei Wang, Yuhua Zhang, Victor C. M. Leung. D2D Big Data: Content Deliveries over Wireless Device-to-Device Sharing in Large Scale Mobile Networks[J]. IEEE Wireless Communications, 2018, 25(1):32-38.

[5]. Adjerid I, Kelley K. Big data in psychology: A framework for research advancement[J]. American Psychologist, 2018, 73(7):899.

[6]. Hemingway H , Asselbergs F W , Danesh J , et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential:[J]. european heart journal, 2018, 39(16):1481-1495.

[7]. Khan S, Shakil K A, Alam M. Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions[J]. Computer Science, 2018, 03(5):107-117.

[8]. X. Wang, Y. Zhang, M. Xu. Development of integrated network platform for heterogeneous agricultural information remote monitoring terminal[J]. Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering, 2017, 33(23):211-218.

[9]. Guo Y , Rao J , Cheng D , et al. iShuffle: Improving Hadoop Performance with Shuffle-on-Write[J]. IEEE Transactions on Parallel and Distributed Systems, 2017, 28(6):1649-1662.

[10]. Zhou X Q , Sun J X , Luo X X . Design and Development of the Mass Image Storage Platform Based on Hadoop[J]. computer knowledge & technology, 2018, 742(264):721-725.

[11]. Chowdam Sreedhar, Nagulapally Kasiviswanath, Pakanti Chenna Reddy. Clustering large datasets using K-means modified inter and intra clustering (KM-I2C) in Hadoop[J]. Journal of Big Data, 2017, 4(1):27.

[12]. Cao Z , Lin J , Wan C , et al. Hadoop-based Framework for Big Data Analysis of Synchronized Harmonics in Active Distribution Network[J]. Iet Generation Transmission & Distribution, 2017, 11(16):3930-3937.