ISSN: 2688-8653

Research on Distributed Visualization System based on Agricultural Big Data

Shiyao Ma[1], Yi Wu[1], Yuanjian Zheng[2], Longling Zhang[1]

[1]School of Date science and technology, Heilongjiang University, Harbin, China, 150080, China

[2]School of Entrepreneurship Education, Heilongjiang University, Harbin, China, 150080, China

Abstract

Relations of production agriculture to survive life issues broad masses of peasants, but also related to socio-economic development. Today, big data analysis and mining covers almost all walks of life but based on farming production applied research still large data processing less . based on this, about the use of distributed visualization system to process agricultural data, to achieve the modernization of agriculture, boost the development of the agricultural industry has a very necessary practical significance and practical value.

Keywords

agricultural big data, distributed , visualization.

1. Introduction

China’s agriculture is faced with resource shortages, increased pollution, domestic production, and international market mismatches. It is urgent to modernize new technologies and invest in agricultural production. Agriculture is one of the sources of massive and multi-source raw data. It is also the most suitable for data. An important field of processing analysis and value rediscovery. This article mainly discusses the use of distributed visualization systems when agricultural big data faces large-scale data processing and display. Real-time processing and utilization of data. Realize the discovery of agricultural data value.

2. Agricultural Big Data Meaning and Prospects

Agricultural big data involves various links such as land preparation, breeding, seed covering, fertilizer, plant protection harvesting, storage and transportation, agricultural product processing, sales, and high animal husbandry production. It is cross-industry and cross-professional data analysis and mining. For food security and food safety, It is of great significance. At the same time, agriculture is an industry with temporal and spatial attributes. Therefore, research-based on different time points and different spatial factors is also necessary. The integration of these data will play an extremely important role in future analysis, mining, and use. an important role. by studying the existing agricultural data. it can improve agricultural production and refinement. intelligent management mechanism in agriculture, methods, and models . can provide guidance to assist agricultural research, government decision-making, agribusiness development opinion.

3. Application of Agricultural Distributed System

In today's society, single computers the ability to handle large amounts of data more and more restricted. The existence of large-scale computing clusters occupies a large space, power consumption, higher maintenance costs a lot of inconveniences. So how as many computer nodes Build a small computing cluster. Reduce the cost of computing as much as possible in

ISSN: 2688-8653

the case of parallel computing performance. Apply the cluster to actual large-scale agricultural data processing. Providing a good data analysis platform for processing big data is the core task of data analysis application research.

The establishment of high-performance distributed computing space and the final data center infrastructure depends on the improvement of computing power. The emergence of big data has brought a qualitative leap to many fields. It also has a new system for computer system structure and storage. to calculate the database. and challenged the organization of large memory data storage and processing methods and new levels of computing systems based on memory. distributed system will also promote agricultural raising it to a new level. Currently. the collected data increasingly agriculture So the establishment of agricultural data centers in areas with better agricultural development is an urgent problem to be solved. The establishment of agricultural data centers aims to more effectively solve the deep problems of agricultural data analysis and agricultural machine learning. Agricultural distributed systems are based on Hadoop Cluster deployment. It is the information support layer of the entire system. From a functional perspective, it is refined into real-time data collection and processing. The execution of data analysis algorithms.

3.1. Real-time Data Collection and Processing

Because different types of sensors, for example, the aerial sensor and ground sensors, data are collected to multiple heterogeneous sources. To avoid data island phenomenon. For different channels of data. The system uses a different message queues to receive.

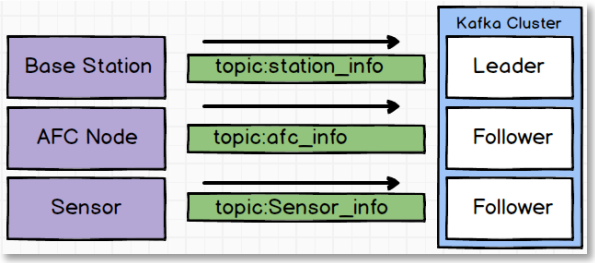

Figure 1. Kafka cluster obtains information from the perception layer

Next, in order to achieve real-time processing of data. The system introduces the Spark-Stream frame. In a certain time window as a unit, from Kafka acquire batch data and the compressed data stream to be delivered to a discrete driver. Drive one job is assigned to a different executor execution. In executor execution of the task, data processing, and filtering processing based on stream, the overall process can be roughly divided into stream Map - Reduce process, wherein Reduce functional data. the need to store the data offline analysis HDFS file system and Mysql database. delivery will require real-time data to demonstrate agricultural Redis message queue.

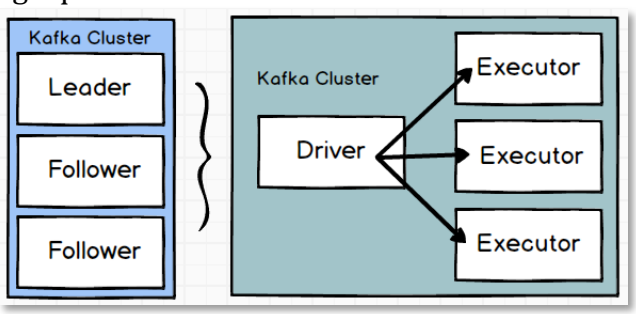

Figure 2. Spark consumes data from the Kafka cluster in a secure time window

ISSN: 2688-8653

3.2. Implementation of Data Analysis Algorithms

Analysis algorithms currently available in the market is mainly composed of two parts: a lightweight data by means of python to assist in the analysis file. Background using Java language development. Based on craft tools and Linux build system SFTP connections. And remote call Python file results. For heavyweight data is often used Spark-milab solution. the framework calls at specified time window Bean component of HDFS file system polling if the new calculation results. then read out the stored Mysql data visualization component requests waiting at any time call. Further. SpringBoot also set a plurality of listeners. when the data requiring real-time display has been calculated. the message immediately by WebSocket sending the page to the visual channel.

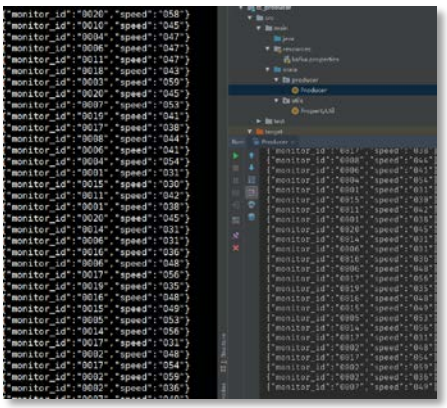

Figure 3. The program sends large amounts of data to the cluster

3.3. Visualization System

Current Visualization technology plays in modern farming production process an important role. Will be transmitted in a real-time distributed system of crop growth data. After the analysis of algorithms. Transferred to the visualization system. Managers can fully observe and monitor crop growth. Can Discover potential hidden dangers in time.

The steps of the visualization system can be roughly divided into first processing the data into static resources and server resources. Combining the CDN content distribution network and Nginx load balancing service. Realizing the reading and saving of data. Use D3.js to make up for the chart display Echarts has distortion in the rendering of large amounts of data. It uses SVG format chart output. The two combine the display data. It provides in-depth interactive data exploration. The layer is based on the WebSocket protocol for instant communication with the business application layer. Real-time data processing results The graph reaches the second level response.

4. Conclusion

With the improvement of people's living standards and the advancement of our technology. Modern technology increasingly plays in the production of agricultural planting stage an important role. Full advantage of the capabilities of modern technology. Committed to improving agricultural crop production efficiency and effectiveness. Promote agriculture healthy and stable economic development. enhance the broad masses of peasants farming technology. promote agricultural productivity, improve efficiency.

Acknowledgements

This project is funded by the School of Entrepreneurship Education of Heilongjiang University. The funding number is 201910212189.

ISSN: 2688-8653

References

[1] Pan Y, Tian Y, Liu X, et al. Urban big data and the development of city intelligence[J]. Engineering, 2016, 2(2): 171-178.

[2] Wei L, Yang S Y. Based on Big Data Technology Analysis on the Mode and Countermeasures of Smart City Construction Operation Management[C]//2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE). IEEE, 2018: 184-186.

[3] Bibri S E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability[J]. Sustainable cities and society, 2018, 38: 230-253.

[4] Yi Liu, Jiawen Peng, and Zhihao Yu. 2018. Big Data Platform Architecture under The Background of Financial Technology: In The Insurance Industry As An Example. In Proceedings of the 2018 International Conference on Big Data Engineering and Technology (BDET 2018). Association for Computing Machinery, New York, NY, USA, 31–35.

[5] Yanjun Z, Xiaodong Y, Yi L, et al. Research on the Construction of Wisdom Auditing Platform Based on Spatio-temporal Big Data [J][J]. Computer and Digital Engineering, 2019, 47(03): 616-619.

[6] Z. Zhao, J. Wang and Y. Liu, "User Electricity Behavior Analysis Based on K-Means Plus Clustering Algorithm," 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), Dalian, China, 2017, pp. 484-487, doi: 10.1109/ICCTEC.2017.00111.