

# Research on Big Data Technology-Based Agricultural Information System

N.F. Xie, X.F. Zhang, W. Sun, X.N. Hao

Agricultural Information Institute, the Chinese Academy of Agricultural Sciences  
Key Laboratory of Agri-information Service Technology, Ministry of Agriculture  
P. R. China

**Abstract**—With the advent of the era of cloud, big data technologies have attracted more and more attention. In the agricultural domain scientists have encountered restrictions and obstacles in the data-extensive analysis domain such as meteorological data processing, bio-genome analysis, food security analysis, because of the huge data set. The paper proposes a big data procession technology hierarchy of agricultural information system from the following aspects: gathering, store, analysis and visualization of agricultural big data. An agricultural information system will be discussed based on large data technology, and it shows that the method is effective.

**Keywords**—big data; agricultural information system; data-extensive analysis; system integration

## I. INTRODUCTION

Agriculture big data is big data concepts, techniques and methods practiced in the agriculture domain. Now, first and foremost farmers have to measure and understand the impact of a huge amount and variety of data which drive overall quality and yield of their fields in which weather data, GPS data, soil specifics, seed, fertilizer, sensor data are being generated a large amount of data each second. Integration and future mining of these data for the development of modern agriculture will play an extremely important role.

Current world governments, social organizations are aware of the big data opportunities and accelerating agricultural big data-based application development. The U.S. government has established a public data website ([www.data.gov](http://www.data.gov)) to provide various types of information collected by the government for the whole society. The system has included 348 agricultural data sets. In May 2013, the leaders of the eight countries (G8) had agreed that an open data sharing platform should be build to cancel the data limitations of agricultural data use. The Climate Corporation in California has used its unique data collection and analysis platform, which gets weather data from 2.5 million collection points every day, and provides many services, such as weather simulation, plant roots and soil structure analysis of massive information on the accident to make a weather risk comprehensive judgment, and also provides crop insurance to farmers. Farmeron Company in Croatia provides data similar to Google Analytics tracking consulting services for farmers around the world to help farmers organize fragmented agricultural production information together with advanced analysis tools such as monitoring, analysis production status of the farm so that it can scientifically helps farmers in agricultural production plans. Japan has begun to use big data for agricultural

production, gathering the indicators of agricultural products through sensors, cameras, various terminals, and then developed cloud service for real-time monitoring, analysis and management.

In China, the research and application of agricultural big data is just beginning. April 2013, Agricultural Information Institute of CAAS and the Network Center of CAS have built a joint agricultural research data center. In 2013, Agricultural Information Institute of CAAS launched agricultural information coalition organization with Academy of Agricultural Universities, and other agricultural enterprises to make full of big data technology in Agricultural domain. June 18, 2013, Shandong Agricultural University established a "Big Data agriculture industry technology innovation strategic alliance" to promote the organization of agricultural research and application of agricultural big data. The agricultural data technology will play important role in agricultural domain and it will bring a revolution for agricultural data use.

The paper is structured into four main sections. Section 2 describes big data processing and analysis technology. In section 3, agricultural information system based on large data technology is presented and described in detail. Finally, Section 4 provides some conclusions and summary remarks.

## II. BIG DATA PROCESSING AND ANALYSIS TECHNOLOGY

To process large data for getting the data value is becoming very import because more and more types of data are produced and the current technology is increasingly unable to handle it so that the data value is burn in the data ocean. How to guarantee smooth operations, and effective management and analysis of data mining useful information that allows decision-makers to do the most correct decisions is now current research focus, which needs big data processing techniques to solve the problems [2-3]. Therefore, a complete set of techniques and tools to analyze and understand big data is necessary for data owner to discover the value in data.

In the entire data processing, the basic processing of big data with traditional data processing is not different because big data is a large volume unstructured data which can not be handled by standard database management systems like DBMS, RDBMS or ORDBMS. In each steps the related tools are being developed with MapReduce parallel processing technology for data processing[4], As showed on Figure 1. The key technologies include the following aspects:

1) Large data gathering: the different of data collection methods is to acquire data generated from the source including

interaction between people and people, people and machine, the machine and the machine such as Web, networking, mobile phones, sensors and so on.

2) Large data storage and management; NoSQL are used for large data storage and management. The data processing mode is completely distributed in various low-cost servers and storage disk so that it helps the web page and a variety of interactive applications quickly process massive amounts of data. It uses a series of distributed technology for real-time analysis of massive data to meet the business needs under big data environment.

3) Big data analysis/mining: Big data technology is about how to analyze and dig out the value for data source. Traditional relational database for large data analysis tasks can not be qualified, because the systems are parallel relational database systems to pursue high degree of data consistency and fault tolerance. MapReduce and Hadoop are as the representative of non-relational data analysis techniques and suitable for large-scale parallel processing. It has achieved significant progress in internet information search and other big data analytics and has become mainstream technology.

4) Large data visualization. It is big data analysis / mining results representation [1]. Generally it refers only to reporting tools & business intelligence analysis product. It is coming a modern branch of descriptive statistics. It involves the creation and study of the visual representation of data, which means that information that has been abstracted in some schematic form, including attributes or variables for the units of information.

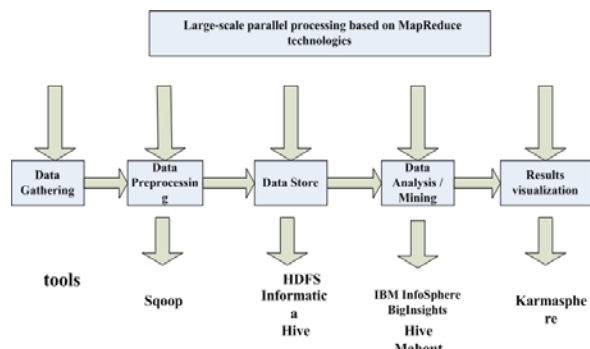


FIGURE I. LARGE DATA PROCESSING FLOW.

As seen from the above flow chart of the processing of large data, MapReduce technology is through the cycle of the large data processing using MapReduce parallel processing techniques to increase the data processing speed and is designed through a lot of low-cost servers for large data parallel processing. It highlights the advantages of scalability and availability, especially for massive structured, semi-structured and unstructured data mixing process [5,7]. It is a programming paradigm that allows for massive scalability across hundreds or thousands of servers in a Hadoop cluster.

MapReduce is a programming model and an associated implementation for processing and generating large data set that is amenable to a broad variety of real-world tasks. Users specify the computation in terms of a map and a reduce function, and the underlying run-time system automatically

parallelizes the computation across large-scale clusters of machines, handles machine failures, and schedules inter-machine communication to make efficient use of the network and disks[10].

### III. AGRICULTURAL INFORMATION SYSTEM DESIGN BASED ON LARGE DATA TECHNOLOGY

The current agricultural information system infrastructure must meet big data processing demand. This requires not only the gradual standardization and upgrading of the business system, and its basic platform facilities should be able to expand and integrate resources for avoiding the construction of a simple computing power and storage capacity requirements, as showed on Figure 2.

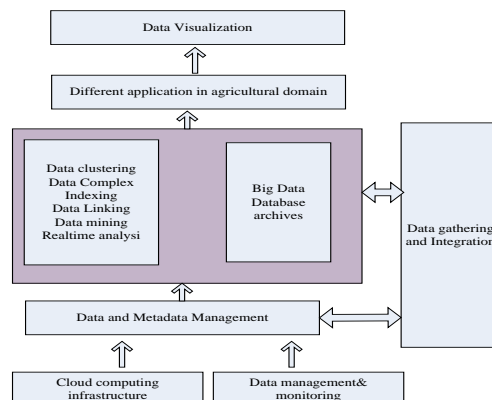


FIGURE II. INFORMATION SYSTEM HIERARCHY BASED ON BIG DATA TECHNOLOGY.

The construction and improvement of infrastructure is considered for sustainable development and the proposed expansion of a comprehensive solution which will enhance the overall efficiency of the existing system to ensure that it will not affect the reliability, availability, and security improvements in the application system. The researchers generally believe that cloud computing technologies are used to build infrastructure and platform including core technologies and services such as cloud computing, virtualization, cloud storage, cloud security, and other key technologies, which will be designed to meet the requirements of better data storage, data protection and more secure, better grading process data. Therefore, the construction of infrastructure and improvement of agricultural information platform should take advantage of cloud computing technology model to make sure that supports the efficiency of data storage, data analysis, data management, and data.

In order to build an agricultural big data analysis application platform, we rely on related data analysis and processing technology based on agricultural big data. The framework of agricultural big data systems uses the technologies of the internet of things to gather data, and use cloud computing technologies for data processing so that the big data system should be built on MapReduce technologies, which will have good expansibility and applicability. Based on the big data technology frame, intelligent decision support system should be developed.

#### IV. AGRICULTURAL INFORMATION SYSTEM DESIGN BASED ON LARGE DATA TECHNOLOGY

We developed an effective domain search engine system, called DSES, which can obtain results by a unified interface to several searches and aggregates them. We used MapReduce to get 6000 distributed documents from Goole, Baidu, and Sogou search engine. In the process, each document will be computed and labled as rank value by the relevance between document and query in parallel, and then use a reduce process to combine them. Mapping and reducing operations are multi-threaded to achieve even greater efficiency in modern multi-core machines, and improves the performance by about 35% over a non-Mapreduce implementation. The figure show the last result from information integration of different search engines.



FIGURE III. THE INTERFACE OF DOMAIN SEARCH ENGINE SYSTEM.

#### V. CONCLUSION

Data has become strategic resources as important as natural resources, human resources, implied great value, and has caused the scientific community and the business community's attention. With the recent rapid data growth, the existing data processing technology is difficult to meet the large demand and data value is very difficult to mine. Therefore, we must find and apply a new data processing technology to mine a data value from agricultural big data. This paper describes how to deal with the flood of agricultural data from the view of the big data technologies:

(1) Deciding that which method is the most suitable for solving a particular data analysis is often critical because each data domain has own data analysis model. So data analysis efficiently is necessary for agricultural data value mining.

(2) The big data technologies are the significant role in the development of agriculture modernization, but agricultural research and applications are required orderly by their importance and the government encourages the research institutes and enterprises whose have extensive data resources and technological advantages to carry out the research on agricultural data collection, storage, transport, mining, research and analysis, application.

(3) In the recent years, information system integration or business integration are concerned too much [8]. Now we must pay attention agricultural data integration in the big data area because the data can bring out its value once the data is gathered and stored in an integrated database.

#### ACKNOWLEDGMENTS

The work is supported by CAAS Agricultural Science and Technology Innovation Program and key projects of Ministry of Agriculture on the cultivation of new varieties of genetically modified organisms (No. 2014ZX0801101B).

#### REFERENCES

- [1] Iliinsky, N. & Steele, J., Designing Data Visualizations [J]: Representing Informational Relationships. O'Reilly Media, 2009.
- [2] Dean, E., Ghemawat S., Big Data: A Revolution That Will Transform How We Live, Work, and Think [J]. Magazine Communications of the ACM - 50th anniversary issue: 1958 - 2008 CACM Homepage, 51(1): 107-113, 2008.
- [3] Herodotou, H., Lim H, H, Luo, G & etc., Starfish: A Self-tuning System for Big Data Analytics[C].5th Biennial Conference on Innovative Data Systems Research (CIDR'11) January 912, 2011, Asilomar, California, USA.
- [4] Manyika, J., Chui, M., Brown, B. & etc., Big data: The next frontier for innovation, competition, and productivity[D], 2011, see <http://www.citeulike.org/group/18242/article/9341321>.
- [5] Abouzeid, A., Bajda-Pawlikowski, K., Abadi, D. J., Rasin, A., and Silberschatz, A., HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads[C]. PVLDB, 2(1), 2009.
- [6] Roger, B., Short, J., & Baru, C., How much information? 2010: Report on enterprise server information [J], University of California, San Diego, Global Information Industry Center, January 2011.
- [7] Dittrich, J., Quiané-Ruiz, J.A., Efficient Big Data Processing in HadoopMapReduce [J]. PVLDB, 2012, 5(12):2014-2015.
- [8] Xie, N.F., Wang, W.S., Research on 3G Technologies-Based Agricultural Information Resource Integration and Service,," CCTA 2009, pp. 114-120, 2010.
- [9] [http://en.wikipedia.org/wiki/Big\\_data](http://en.wikipedia.org/wiki/Big_data).
- [10] Dean, J., Ghemawat S., MapReduce: a flexible data processing tool[J]. Commun. ACM 53(1): 72-77, 2010.