The architecture design for big data application system in apple industry

Yandong Bi [1], Peng Wang [1], Zhijun Wang[1] and Shuhan Cheng [1*]

[1] College of Information Science and Engineering, Shandong Agricultural University, Tai'an, Shandong, 271018, China

[*]Corresponding author's e-mail: shcheng@sdau.edu.cn

Abstract. Aiming at the problems of complex data sources, poor management level and blocked market information in the orchard, this paper constructs the index system of apple big data and designs the functional framework and technical framework. Through collecting heterogeneous data sources such as fruit tree cultivation, meteorology, soil, farming work, diseases and pests, market circulation and fruit quality for mining analysis, the big data application system of apple industry is established, which provides apple industry development planning, precise irrigation and fertilization of orchards, apple quality demand analysis, domestic and international marketing risk early warning and other functions for fruit farmers, agricultural organizations and governments. At present, the system is reliable and practical, and plays an important role in big data of fruit industry in China.

1. Introduction

Chinese agricultural big data started late. Since the founding of the first "strategic alliance for technological innovation of agricultural big data industry" in 2013 in China, many provinces and municipalities have established agricultural big data platforms. In 2015, China carried out a big data pilot project for the single agricultural product. In 2020, the government data will be open to the public. So far, domestic big data industry has developed rapidly. In view of the high cost of data acquisition, Gu Geqi[1] proposed an agricultural big data collection platform based crowdsourcing. For big data security risks, Qin Shasha[2] discussed data protection issues. Meng Xiangbao[3] proposed the architecture and platform construction of application system for agricultural big data; Tian Cheng[4] proposed a micro-agricultural big data platform based on Raspberry Pi 2. Xie Nengfu[5] put forward a knowledge fusion framework based on agricultural ontology and fusion rules. Although some outcomes have been gained, we still lag far behind the developed countries such as Europe and America. To further improve demestic big data technology, we can learn from the successful cases of other countries. For example, the United States developed farmland management softwares for agricultural big data, such as Fram Sight, Climate Pro, Scirpts and Field 360. Abrahamsson[6] used Raspberry Pi hardware to build an energy-efficient cloud computing cluster, which improves the low computing power of Raspberry Pi and is suitable for severe environments. Waller[7] proposed the application of big data analysis technology in the supply chain. Based on the Hadoop framework, Mohammed[8] discovered crop cultivation patterns under different agriclimate. Demirkan[9] proposed a conceptual framework for DSS in cloud and demonstrated the opportunities and challenges of engineering service oriented DSS in cloud. At present, there are few large-scale data architectures for the whole industry chain of apple. According to the complexity and regularity of production process of apple, this paper designs a big data application architecture for apple industry, in order to provide a reference for the development of agricultural big data in China.

2. Apple industry big data application architecture

At present, apple industry are faced with many problems such as increase of production without relevant increase of benefit, increase of efficiency without relevant increase of income, the weak response to domestic and foreign markets, insufficient understanding of the diversified needs of consumers and increasingly prominent problems of resources and environment. It is an urgent need to apply the Internet of Things, big data and artificial intelligence to establish a multi-source data-aware, content understanding, deep learning, deductive reasoning integrated technology system and application platform, so that information technology is widely used in the industry chain, which lays the foundation for agriculture 4.0.

Before production, high standard farmland construction, farmland environmental monitoring, agricultural materials supervision, and market forecast analysis are needed. In production, intelligent agriculture is needed to promote the integration of agricultural machinery and agronomic intelligence so as to save costs, increase efficiency and improve land productivity, labor productivity and resource utilization. After production, we need to mediate the relationship between agricultural product e-commerces, cold chain logistics, agricultural product quality safety monitoring and warning. At the same time, information feedback and machine learning is a must to improve the service level of agricultural big data.

According to the related elements of the smart orchard, we establish demand analysis. From the viewpoint of system and according to the top design principles, the apple big data SMART(Specific, Measurable, Attainable, Relevant, Time-based) application architecture is put forward from five aspects of service, management, application, resource and technology. The architecture is shown in Figure 1. In the figure, technology and resources are the basic (input) layers of agricultural big data application; analysis processing and application platform is the most direct output of agricultural big data application, which is defined as application layer; management is to provide automation, integration and intelligent management in the planning and construction of apple industry projects and the operation and maintenance of various application systems; service refers to the provision of apple industry public services and professional technical services for the vast number of fruit farmers, related enterprises and organizations; management and service constitute the service layer of apple big data service platform architecture.

The application layer makes full use of cloud computing, pattern recognition, artificial intelligence and other intelligent computing and information processing technologies to analyze and process massive data and information, and dynamically optimize production, management and marketing programs, which is the core of data processing. The service, management, application, resource and technical requirements of apple big data service platform are further explained below.

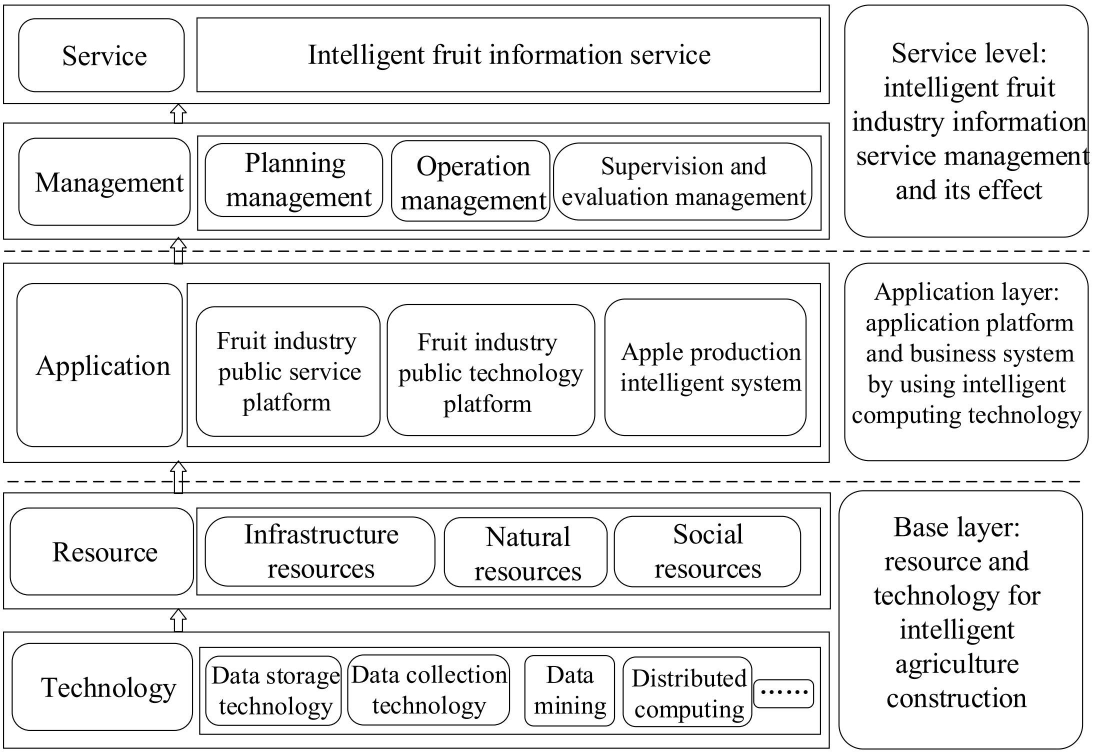

Figure 1. Agricultural big data application structure system.

2.1Service

Service is the ultimate goal of apple big data research and application. Fruit farmers, agricultural corporations, farmer cooperative organization, related enterprises(production, processing, storage, transportation, sales) are the main actors of the apple industry; the government is the main provider of public services for the apple industry, providing infrastructure construction, technology promotion, informationization, policies, laws and regulations of public services, including comprehensive, systematic and integrated services before, during and after apple production. All actors must provide relevant data and information in time and actively participate in the construction of apple big data service platform to solve the problems of data sharing and deep utilization.

2.2Management

Management refers to the management functions that the government and the main actors of apple industry should perform in the process of apple industry activities, from the project planning construction to the maintenance and supervision management of the application system. The whole process needs the participation of the government and all actors. The management of smart agriculture should be based on macro-management, which mainly includes three parts: planning management, operation management, supervision and evaluation management. Planning management mainly refers to scientific and reasonable planning, organization and coordination in the major projects of apple industry in manpower, resources, funds and other aspects on the basis of relevant industrial planning, policies and regulations; operation management covers fruit production, processing, circulation, safety and other fields, which is embodied in 3 aspects: management mode, revenue mode and operation effectiveness. The supervision and evaluation management is mainly to supervise and evaluate the effect of planning and operation management.

2.3Application

Application refers to the design, development and use of apple big data application system and management platform, which provides comprehensive support for apple big data management and service. According to the main sources of apple big data, the application areas can be summarized as follows:

2.3.1Apple production management. Data collected through agricultural Internet of Things (IOT) technology in the apple production process is used to analyze to provide "accurate" agricultural formula and agronomic measures, "intelligent" management decision-making and facilities control. Agricultural big data has played an important role in the digital orchard management, such as fruit tree growth monitoring, orchard pest monitoring and real-time identification, real-time provision of pest control programs, apple production professional guidance, and fruit quality and safety traceability[10].

2.3.2Fruit safety management. Fruit safety management involves all aspects of apple food chain, such as environment of apple production area, pre-production, production and post-production, industry chain management, storage and processing, market circulation, logistics, supply chain and traceability system. Through the analysis and processing of apple quality and safety supervision information, the early warning of fruit safety risks and emergency management of quality safety can be realized[11]. It is necessary to establish a credit evaluation system for the main actors of apple industry, including fruit farmers, online retailers, fruit processing enterprises, so as to make the whole industry chain process more transparent.

2.3.3Ecological environment management. Apple ecological environment includes soil, atmosphere, water quality, meteorology, pollution, disasters and so on. It is necessary to monitor these environmental factors comprehensively and manage them accurately in order to explore the effective way of "double control and double drop" of chemical fertilizer and pesticide in orchard through big data technology. According to historical data and disaster data, disaster early warning mechanism and agricultural insurance services should be provided on demand to effectively reduce the huge losses of fruit farmers due to disasters.

2.3.4Planting area map management. The information of orchards in various planting areas is collected through GIS technology, remote sensing technology and camera equipment, and the real-time status of soil and plant nutrition is extracted according to the pictures. Based on the diagnosis method of plant and soil deficiency, the method of soil testing and formula fertilization and the expert knowledge base, the instructions are given to the agricultural machinery so as to carry out the corresponding variable fertilization, irrigation, sowing and other farming operations.

2.3.5Fruit supply and demand management. Fruit supply and demand management includes supply and demand analysis. The agricultural large-scale data analysis technology was used to classify the high-quality varieties of the national production areas, and to build brand marketing strategy and to provide advertising for high-quality producing area, and to establish a quality evaluation system. According to the fruit trading data, e-commerce trading evaluation module (need to check whether there is a malicious attack or scalping operation) and other ways to analyze the demand of various varieties, personalized recommendation for consumers and appropriate guidance to fruit farmers to select varieties of planting. In order to control the price changes at the macro level, the users should register their annual cultivar. When supply exceeds demand in the market circulation, early warning of this variety planting should be provided for fruit farmers.

2.3.6Orchard equipment management. It is necessary to provide intelligent management and application of equipment and facilities in the monitoring, remote diagnosis and service dispatch under the working conditions in the orchard. By utilizing some devices that collect environmental information such as orchard meteorology and soil, monitor physiological information such as fruit tree nutrition and fruit quality, detect plant biomass and properties, and apply variable pesticide, the accurate operation technology system of orchard management informationization can be established and can provide the orchard information service, accurate operation technology and equipment integration application demonstration.

2.3.7Apple scientific research big data rescource management. The big data produced from apple scientific research include spatial and ground remote sensing data, gene map, large-scale sequencing, apple genomic data, and a large number of biological experiment data. The intelligent analysis of apple scientific research data can better guide apple production, processing and circulation. In the above applications, the fruit production process, the fruit resources, the ecological environment, the fruit quality and safety, and the monitoring and forecasting of fruit market circulation are the key application direction.

2.4Resource

Resources are divided into natural resources and social resources. The main consideration is natural resources, infrastructure and information resources. Natural resources include land resources, water resources and atmospheric resources. Infrastructure resources here refer specifically to the network infrastructure including wired transmission networks, wireless sensor networks, mobile Internet, and various information terminals, which is the communication infrastructure foundation of apple big data applications. Information resources refer to languages, texts, audios, videos, images, charts, figures, and information collected by sensors in the process of apple production. At present, there are many problems in apple production and management activities, such as scattered resources and inconsistent standards, which makes it difficult to share and exchange information. Therefore, it is necessary to manage the metadata effectively, unify the exchange standards, optimize the exchange process constantly, and provide a powerful data information system for the application of apple big data.

2.5Technology

With the rapid development of information system, not only a large amount of information is required to be stored reliably, but also a large number of users need to rapidly access the required data. Big data technology is a whole without unified solution. Related technology involves data transmission, storage, calculation, mining, display and development platform of six parts. HDFS[12] in Hadoop has the advantage of massive data storage and management, Spark[13] is a big data offline processing technology based on memory computing, whose processing speed is better than Hadoop; Storm[14] is a real-time data processing technology based on stream computing. In view of the characteristics of apple big data and the functional features of various storage applications, apple big data application technology architecture can choose HDFS as the agricultural large data storage and management infrastructure, and choose Spark or Storm architecture to analyze large-scale data. The technical architecture of Hadoop is shown in Figure 2.

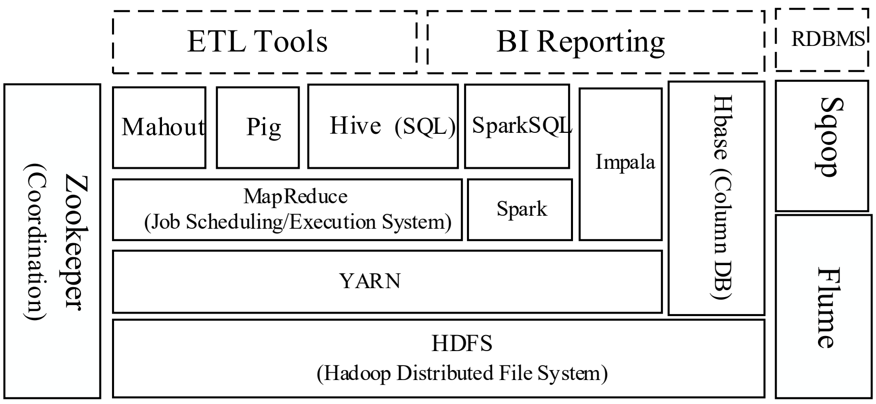

Figure 2. The Hadoop technological structure for apple big data application.

3. The technological design for big data intelligent analysis platform in apple industry

One of the important goals of big data analysis is to discover the relevance of data. The low-value density characteristics of agricultural data require in-depth comparison, purification and transformation of large-scale data, in order to find its inherent significance and mutual links. The platform designed in this study can be used in the intelligent digital orchard, and provides intelligent decision support for fruit farmers, related enterprises and departments.

3.1The overall structure for platform

With the help of advanced and mature large-scale data storage and processing technology, an intelligent analysis platform for apple big data is constructed to form an authoritative unified view of fruit industry information. This research designs the overall architecture of apple big data intelligent analysis platform from three aspects: IAAS (Infrastructure as a Service), PAAS (Platform as a Service) and SAAS (Software as a Service)[4], as shown in Figure 3.

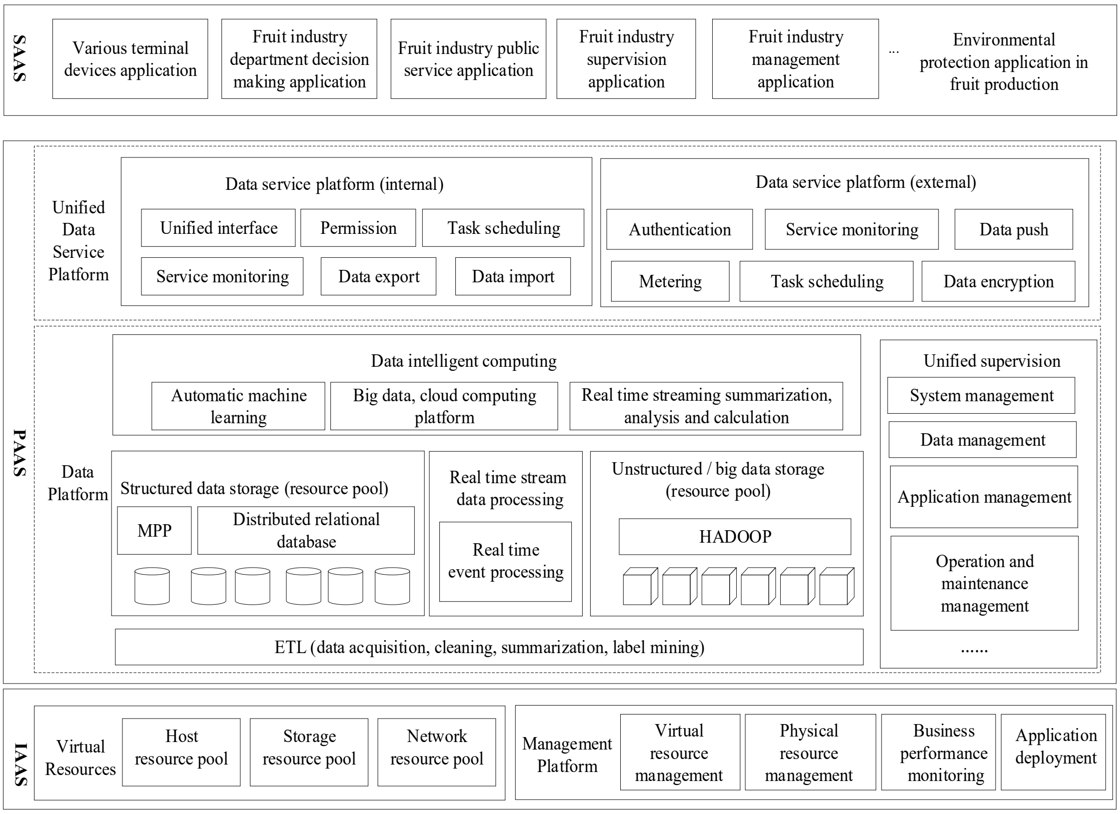

Figure 3. The overall structure of intelligent analysis platform for apple big data.

3.2The implementation of platform technology

3.2.1Technological structure. Apple big data intelligent analysis platform is composed of six parts: integrated platform, ETL(Extract-Transform-Load) Tool, design tool, runtime, preset application and BI (Business Intelligence) portal. It provides end-to-end BI services for related enterprises. The technical framework is shown in Figure 4.

The integrated platform provides basic system services and operation framework to unifiy the management of various BI tools and analysis models, and resources. ETL tools can complete the integration of heterogeneous data, help to construct data warehouse, and form the subject databases in various fields. Design tools provide flexible query, multi-dimensional analysis, indicators tools, management cockpit, intelligent reporting, and map analysis tools to achieve the definition and release of various analysis models. The runtime is used to analyze the design model and monitor the running state of the whole model. Preset applications provide users with a number of analysis, evaluation, early warning, prediction, optimization models for a reference. BI portal integrates various application system resources, data resources, information resources into a unified platform. According to the different characteristics and roles of each user, personalized application interfaces is formed.

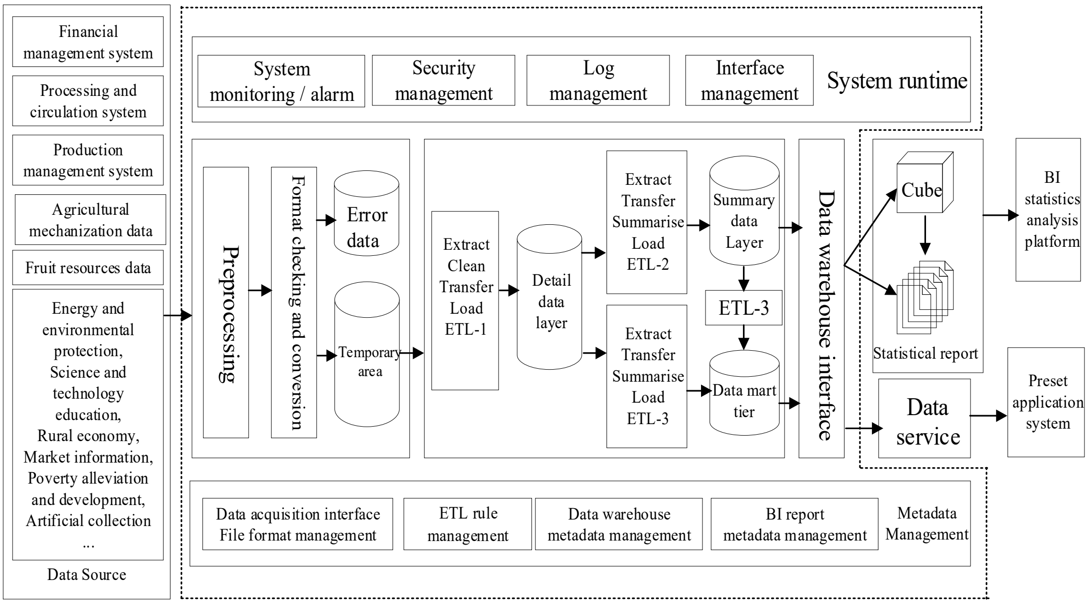

Figure 4. The technological structure for agricultural big data intelligent analysis platform.

3.2.2Data collection analysis. The basic core of the platform is to collect data from various agricultural data sources, organize and form subject databases of apple industry in various fields. The ETL tool, collecting core data, is redeveloped based on the open source project kettle (ETL tool written in pure Java). The ETL tool is efficient in data extraction, stable in performance, and adopts plug-in framework to develop corresponding data acquisition and collation plug-ins for different subjects of data source requirements. The client side realizes non-programming, visual data acquisition functions. The data acquisition process is shown in Figure 4. The data collected from various data sources are stored in the temporary storage area after format checking and conversion. The temporary storage area data is extracted and processed after three levels ETL to form the subject databases in various fields. In the design of subject database, according to different application requirements, three levels of data service layer are designed, which are detail data layer, summary data layer and application market layer. According to the preset ETL extracting rules, the first-level ETL extracts and organizes the data from the temporary storage area into the detail data layer. Second-level ETL extracts, collates, and analyzes detail data into a business-oriented statistical dataset, that is, the summary data layer. The third level ETL extracts data from the detail data layer and the summary data layer into data mart tier to form multi-dimensional data cubes of various subjects.

3.2.3Implementation of apple big data system. The BI analysis platform and portal are implemented by Java. The data mining module is implemented by Python, which is flexible in Hadoop. Firstly, heterogeneous data sources are collected into Hadoop. The data from the relational databases are uploaded to Hadoop by Sqoop, unstructured data are stored to Hbase database, and semi-structured data is uploaded to HDFS. Secondly, ETL technology is used to clean, extract and integrate data, then, the integrated data is analyzed and processed through machine learning algorithm library such as

Mahout and Spark. Finally, the result tables of analysis and mining are stored in the relational database for BI system's visibility of strategic decision. BI system provides apple industry development planning, precise irrigation and fertilization of orchards, apple quality demand analysis, domestic and international marketing risk early warning and other functions for fruit farmers, agricultural organizations and governments.

4. Conclusion

Apple big data service is a new research field, and big data platform is rare. Apple industry has the characteristics of wide distribution of data sources, strong timeliness, diverse types, complex structure and difficult access, which determines the complexity, diversity and challenges of apple big data research and application. This study focuses on the analysis and construction of the apple industry big data application architecture, and the functional design and technical implementation of the apple big data intelligent analysis platform, which has a certain significance for the integration, mining and use of large amounts of data generated in apple production, management, processing and circulation.

References

[1] Gu, G.Q., Li, J. (2018) Construction of agricultural data collection platforms based crowdsourcing. Jiangsu Agricultural Sciences, 46:191-194.
[2] Qin, S.S. (2016) Discussion on the protection of big data in agriculture. Jiangsu Agricultural Mechanization, 6:51-52.
[3] Meng, X.B., Xie, Q.B., Liu, H.F., Yang, X.Y. (2014) Architecture and platform construction of big data application in agriculture. Guangdong Agricultural Sciences, 41:173-178.
[4] Tian, C., Lu, S.K. (2017) Feasibility study on micro-agricultural big data platform based on Raspberry Pi 2. Jiangsu Agricultural Sciences, 45:202-204.
[5] Xie, N.F. (2012) Research on agricultural ontology and fusion rules based knowledge fusion framework. Agricultural Science & Technology, 13:2638-2641.
[6] Abrahamsson, P., Helmer, S., Phaphoom, N., Nicolodi, L., Preda, N., Miori, L., et al. (2014) Affordable and Energy-Efficient Cloud Computing Clusters: The Bolzano Raspberry Pi Cloud Cluster Experiment. In:IEEE, International Conference on Cloud Computing Technology and Science. Singapore. pp.170-175.
[7] Waller, M.A., Fawcett, S.E. (2013) Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics, 34:77-84.
[8] Mohammed, A.H., Gadallah, A.M., Hefny, H.A., Hazman, M. (2018) Fuzzy based approach for discovering crops plantation knowledge from huge agro-climatic data respecting climate changes. Computing, 100: 689-713.
[9] Demirkan, H., Delen, D. (2013) Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud. Decision Support Systems, 55:412-421.
[10] Jiang, M., Wang, J., Wang, Z.J., Cheng, S.H. (2015) Research and integration of decision support system for apple tree diseases and pests. Shandong Agricultural Sciences, 47:129-133.
[11] Wen, F.J. (2013) Big data in agriculture and new opportunities for development. China Rural Science & Technology, 10: 14-14.
[12] Qin, X.P., Wang, H.J., Du, X.Y., Wang, S. (2012) Big data analysis—competition and symbiosis of rdbms and mapreduce. Journal of Software, 23: 32-45.
[13] Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I. (2010) Spark: cluster computing with working sets. In: Proceedings of 2nd Usenix Conference on Hot Topics in Cloud Computing. Boston. pp.10-10.

IOP Conf. Series: Journal of Physics: Conf. Series **1168** (2019) 032075 doi:10.1088/1742-6596/1168/3/032075

[14] Leibiusky, J., Eisbruch, G., Simonassi, D. (2012) Getting Started with Storm. O'Reilly Media Inc., The United States.