

Review article

Practical use of open data to construct regional big data center for Hanwoo farm

Chan-Woo Lee^{1†}, Dae-Hyun Kim^{3†}, Hee-Jin Kim⁴, Dan-Il Kim^{5,6*}, Yoon-Seok Lee^{1,2*}

¹ Department of Biotechnology, Hankyong National University, Gyeonggi-do, Anseong, 17579, Republic of Korea

² Center for Genetic Information, Hankyong National University, Gyeonggi-do, Anseong, 17579, Republic of Korea

³ Livestock Research Institute, Gyeongsangbuk-do, Yeongju, 36052, Korea

⁴ uLikeKorea Co., Inc. Seoul, 05836, Korea

⁵ Institute of Green-Bio Science and Technology, Seoul National University, Gangwon-do, Pyeongchang, 25354, Korea

⁶ College of Veterinary Medicine, Seoul National University, Seoul, 08826, Korea

***Corresponding authors :** Yoon Seok Lee, Tel: 031-670-5332, Fax: 031-670-5339, E-mail: yoonseok95@hknu.ac.kr
Dan Il Kim, Tel: 033-339-6102, Fax: 033-339-6199, E-mail: danilkim@snu.ac.kr

[†] These authors contributed equally to this work.

ABSTRACT

The purpose of this study was to propose a plan for constructing a regional Hanwoo big data center to provide an integrated management system of feeding, reproduction, and genetic improvement using biological and environmental data. The regional big data center for the Hanwoo industry was constructed as described below. First, Hanwoo digital data were collected from the ICT (Information and Communications Technology) equipment in each farm, ICT companies, and related organizations. Thereafter, the collected data were subjected to file storage real-time using agent, agentless, FTP (File Transfer Protocol), and Open API (Open Application Programming Interface) methods. Finally, to monitor and control disease and reproduction, a Hanwoo big data platform based on AI (Artificial Intelligence) was established. Regional Hanwoo big data center can provide services to users according to a data connection plan appropriate to the situation in real-time Hanwoo public to open big data platform. Therefore, the regional Hanwoo big data center could have positive effects, such as maximizing management efficiency of farms by reducing livestock losses, reducing total feeding cost, and reproduction improvement on farms. Moreover, employment opportunities can be created by recruiting labor related to the big data industry through a public to open data platform. Notably, Hanwoo big data platform is expected to expand to other industries and new markets for services.

Key words: Artificial Intelligence, Big data, Information and Communications Technology, Hanwoo, Open data

Introduction

According to a recent trend report of the Hanwoo industry in Gyeongsangbuk-do, the total number of Hanwoo farms has decreased from 20,268 in 2017 to 19,158 in 2020. However, the total number of Hanwoo has increased from 615,929 in 2017 to 694,855 in 2020 (KOSTAT, 2020a). Based on the change in the number of farms, there are fewer than 50 Hanwoo farms, which decreased from 16,665 in 2017 to 15,115 in 2020, and farms raising more than 50 Hanwoo increased from 3,610 in 2017 to 4,043 in 2020 (KOSTAT, 2020b). Such finding indicates the performance of intensive animal farming on limited land. Intensive animal farming, also known as industrial livestock, is a type of agriculture with a higher level of input and output per animal raised on limited land. Owing to the aging of farming population, there has recently been a demand for the

management of intensive farming using ICT (Information and Communications Technology) equipment to achieve precise management. Therefore, the Korean government has supported the allocation of ICT equipment and services for precise management through a local support project, which aims to supply ICT equipment to 5,750 farms by 2022. Owing to the increasing use of ICT equipment and services, the amount of data produced from farms is rapidly increasing (Cheon DW et al., 2016). Thus, a big data center is urgently needed to enable effective utilization of such data to achieve precise management.

Feeding, water, carcass traits, disease, physiological temperature, and activity data were collected from ICT equipment on farms. These data are then automatically converted into digital data and stored on the farm's PC and ICT equipment company server. When these data are accumulated in a storage equipment, big data should be stored and managed separately.

A trend has been identified in the big data center industry abroad (Kim JH et al., 2014, Jang YJ and Kim TW, 2019). In the United States, the 'Big Data R&D Strategic Plan' was established (Jeong YC and Han EY, 2014). Accordingly, territories in the US were divided into the West (medical), Midwest (agricultural and food), South (health care and manufacturing), and Northeast (energy, education, climate change), and a 'Big Data Regional Innovation Hub' was established. This hub is found in more than 250 organizations in each region and is used to solve social problems (Son CM et al., 2019). In the United Kingdom, the 'Data Strategy Committee' was established under the government to strengthen access to data. Accordingly, new jobs and start-up industries related to big data technology such as manufacturing, distribution, and finance were created, which yielded an economic profit of 216 billion pounds (an estimated 350 trillion won) (National IT Industry Promotion Agency, 2014). The use of big data was also implemented in Japan at the IT Strategy Headquarters in 2012 to help the country become a society that uses IT (Information technology), such as machine to machine (M2M), big data infrastructure, and create new industries through big data (Korea Internet & Security Agency, 2014). In China, the 'Action to promote big data development' plan was announced to utilize the big data industry as a new growth foundation for national development. As a result, China government is investing the big data industry that raise market scale from 280 billion yuan (an estimated 47.7 trillion won) in 2015 to 1.01 trillion yuan (an estimated 172.08 trillion won) in 2020 (Korea-China Science & Technology Cooperation Center, 2018).

Currently, a gap exists in the technical level of big data between advanced countries and South Korea. In particular, the indicator of the convergence technique level is lower than that of the advanced country level (Jang YJ and Kim TW, 2019). According to data from the ITSA (Korea Information Technology Service Industry Association) in 2018, the centralization of the metropolitan area is a serious problem because 61% of domestic big data centers are located in Seoul and Gyeonggi-Incheon, 11% in Chungcheong and Gyeongsangnam-do, and 5.6% in Gangwon-do, Jeolla-do, and Gyeongsangbuk-do (Song JH et al., 2018). As the big data infrastructure is concentrated in the metropolitan area, regions outside of the metropolitan area face disadvantageous competition in the big data industry. In particular, the most of livestock industry is located in the provinces, and the competition in the big data industry is even more subsided. The livestock industry's data share in the big data sector is 1.39% (Open Government Data Portal, 2020), which is an insignificant in the big data industry. To address these problems, the Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Education, Promotion & Information Service in Food, Agriculture, Forestry & Fisheries (EPIS) collects data while operating a big data portal, however, most of the data are used to analyze consumer patterns to provide consumer-oriented customized purchase information services. As a result, the actual utilization in farms is very low.

Most companies that supply livestock management systems to Hanwoo farms are small, thus, the services that can be provided to farms using collected data with the ICT equipment are limited. Accordingly, integrated management system that enables Hanwoo farms to integrate and use data on breeding, improvement, and disease is inadequate. Therefore, it is necessary to establish a smart livestock model and an integrated management system for each species of livestock specialized in the region to achieve more precise Hanwoo management while improving farm management convenience. In addition, to strengthen the competitiveness of Hanwoo farms, a smart farm-based high-tech environment should be developed, and the integrated collection system of ICT data and management data for Hanwoo farms should be prepared.

Therefore, in this study, we suggest a method to collect and connect data related to breeding, improvement, disease, and environmental management in real-time for Hanwoo farms using ICT equipment in Gyeongsangbuk-do and propose a model for constructing an integrated management system to establish the 'Regional Hanwoo Big Data center'.

Status of ICT equipment in regional Hanwoo industry

Status of Hanwoo ICT equipment in Gyeongsangbuk-do province

In this study, we investigated information of Hanwoo ICT equipment in Gyeongsangbuk-do province through a literature search and field survey. The information of ICT equipment related to breeding, disease, and reproduction management in Hanwoo farm were shown in Table 1. The domestic equipment were found to mainly collect temperature and activity data for disease and breeding management. International ICT equipment are similar to the domestic ICT equipment, temperature and activity data for breeding and disease management were collected. Especially, due to the geographical characteristics of grazing, GPS location data were also collected by international ICT equipment.

Through a data survey, three of the thirty-six ICT equipment companies that are distributing ICT equipment in Gyeongsangbuk-do were selected, and the data collected by ICT equipment companies and the number of equipment deliveries in Gyeongsangbuk-do. The types of Hanwoo ICT equipment supplied including automated individual feeding machines, rumen activity analyzers, estrus detectors, and calving sensors. The ICT equipment in Gyeongsangbuk-do were investigated to collect data, such as activity level, weight, rumen activity level, tail movement pattern, and disease status of Hanwoo.

Table 1. The information collected from ICT device in Gyeongbuk province.

Improvement	Reproduction	Data		Product name
		Breeding	Disease	
Breeding value	Estrus, Calving, Pregnancy, Grade, Activity level, bloodline	Distribution, Price, Individual number, Feed and Water consumption, carcass grade	Temperature, Rumen activity, Rumen temperature	Calfrail (Automated individual feeding)
				HR-Tag
	GPS Location, Calving, Estrus, Activity level, Rumen activity, temperature	Feed and Water consumption, Individual number	Rumen temperature, Visual data, Rumen activity level, Sound, Temperature,	ENGs wireless estrous sensor Moocall calving sensor

The information needs of farmers in Hanwoo industry will only increase as they have to make more and more complex decisions on how to increase productivity performance. so, we collected farm's opinions to know whether ICT equipment has effectively contributed to breeding, disease, and reproduction management. The farm's opinion is shown in Figure 1. As shown in Figure 1, Farmers have positive thought about ICT equipment by words of 'ICT', 'Positive', 'Assistance business', 'Helpful'. And the opinions of farmers on the establishment of a big data center for Hanwoo were positive. In fact, words such as 'Positive', 'Disease management', 'Breeding management', 'Feeding management', and 'Integrated management', were frequently used by the farmers. Therefore, most Hanwoo farmers have a positive outlook regarding the establishment of a big data center for Hanwoo. Moreover, these farmers need a system that can manage disease, breeding, and production in an integrated manner (Fig 1).



Journal of Animal Breeding and Genomics

Journal of Animal Breeding and Genomics

Journal of Animal Breeding and Genomics

Journal of Animal Breeding and Genomics

Journal of Animal Breeding and Genomics

Journal of Animal Breeding and Genomics

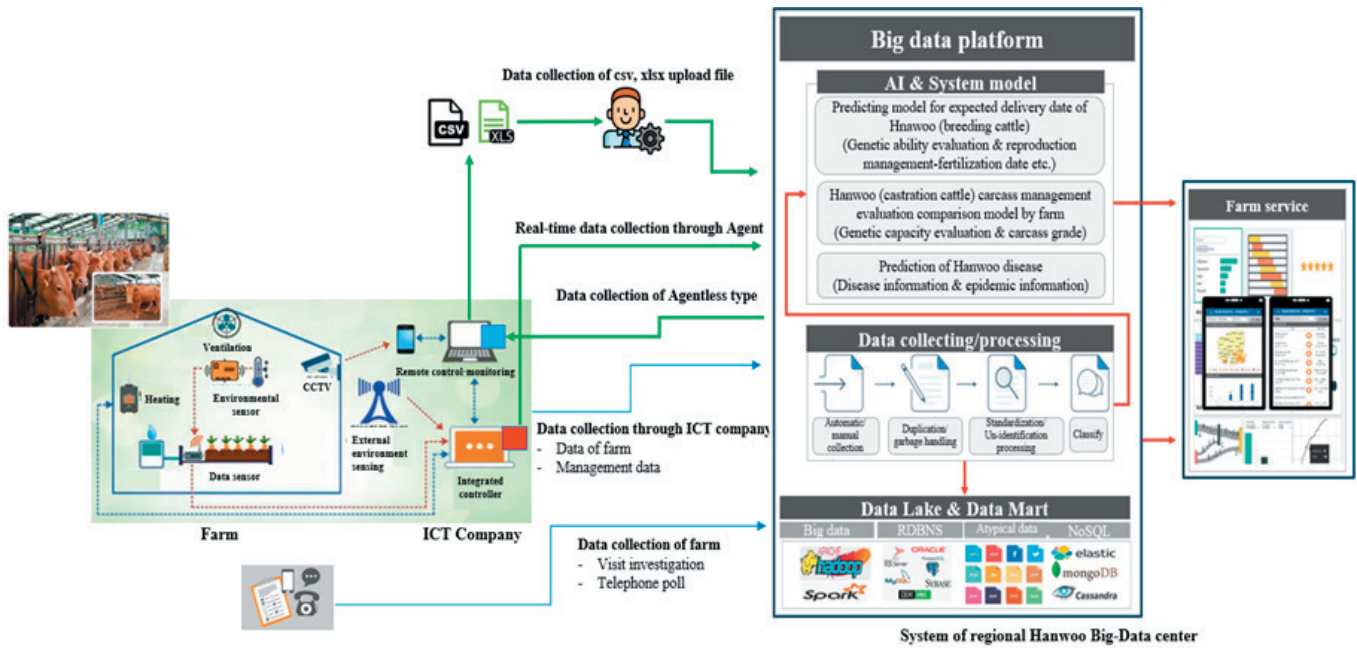


Figure 2. Data integrated system of the farms and ICT companies

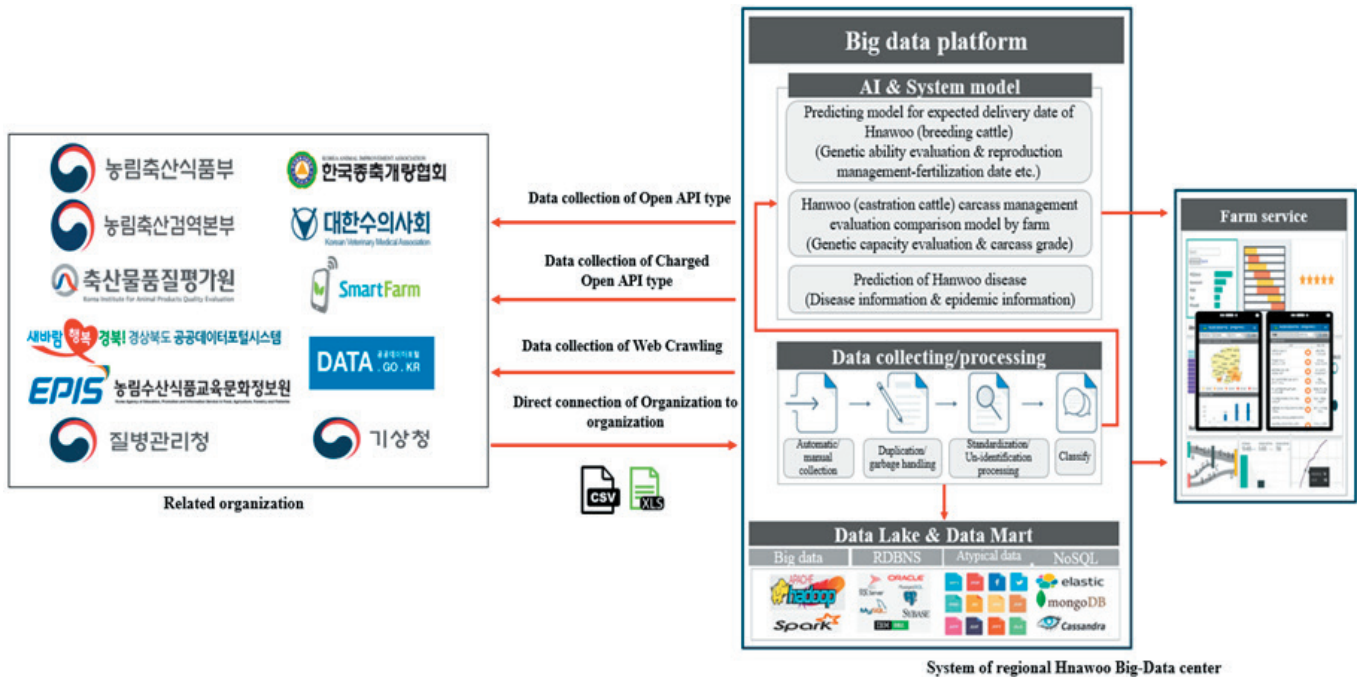


Figure 3. Data integrated system of the related organization

Although the IT systems owned by ICT equipment companies and related organizations differ slightly in scale and complexity, standard systems for data connection between systems commonly use the agent and agentless. Agent is a method of receiving data by installing agent software for obtaining data on the PC of a Hanwoo farm or the ICT equipment companies, agentless is a method of bringing data directly from the data provider's server in an environment where agent cannot be installed. The regional Hanwoo big data center establishes a system by installing agent at ICT equipment companies and Hanwoo data to the Hanwoo big data integrated system. As the data form of related organizations are different for each organization, regional Hanwoo big data center provides agent software to related organizations and connects the server of regional Hanwoo big data center and related organizations. Original data are stored in the Data Lake (DL) and metadata are stored in the data catalog. Thereafter, a system is established wherein the necessary data is retrieved via search of the data catalog.

The collected Hanwoo data is delivered to the user through the steps of 'collection'-'storage'-'pre-processing'-'analysis'-'visualization'. 'Pre-processing' involves processing the collected data into a form suitable for analysis; this is the most important step in data analysis because different results can be derived based on who processes the collected data. The collected original data is stored in the DL through minimal processing in the 'collection' step. To access the data stored in the DL, a data catalog must be provided, and the data retrieved through the data catalog are processed again and stored in the Data Mart (DM). The data stored in the DM are used for analysis and management according to the purpose of the user (Fig 4).

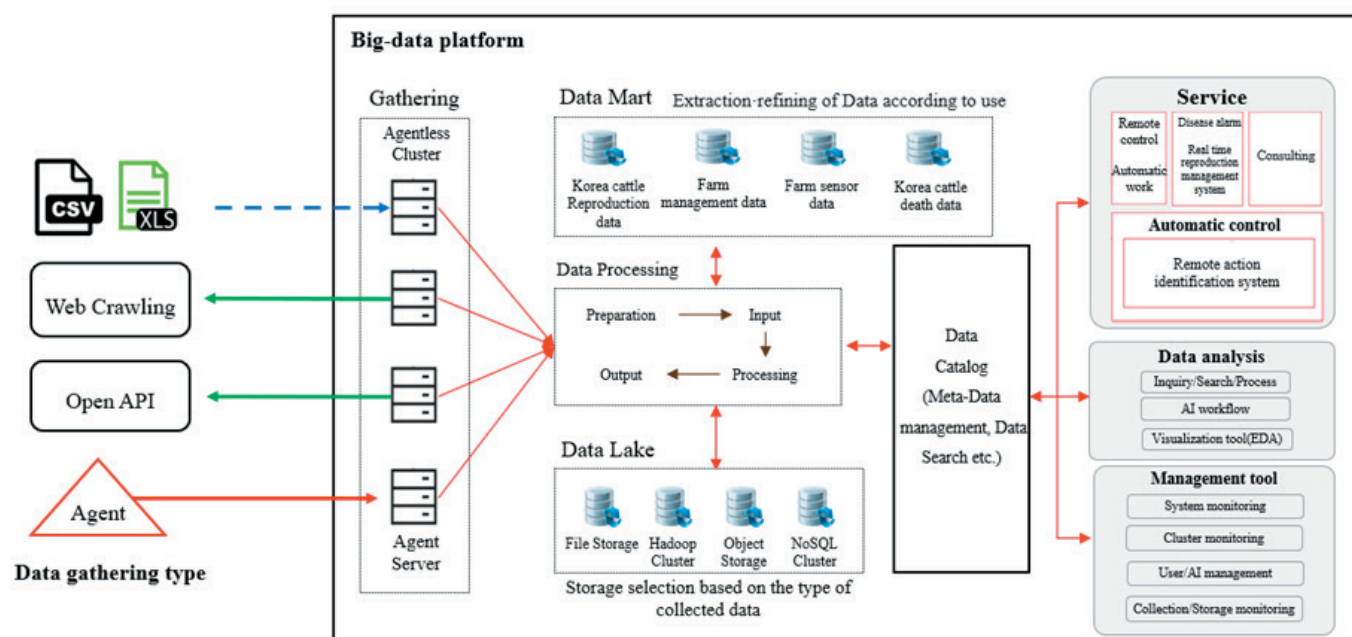


Figure 4. Hanwoo data processing in big data center

Database integration and management system for Hanwoo digital data

Establishment of infrastructure for Hanwoo big data platform

The establishment category of the Hanwoo big data platform can be divided into data collection, processing, data analysis using artificial intelligence (AI), tools, services needed for analysis, and storage infrastructure that can store data. Agent and agentless were used for data collection and processing, and R or Python was used for data analysis. Management tools and services provide tools for managing users, data, resources, and analytical environments for data analysis and system operations. Finally, the hardware for each area is configured as an independent cluster, and the storage is established to ensure that the service is not interrupted even when scale-up and scale-out, which causes system expansion by adding hardware, disk, and memory, are performed.

Development of Hanwoo big data analysis and AI learning system

The software for data analysis using AI should be composed of open source-based software, except for security solutions and operating systems. In addition, when inquiring the data collected for analysis, various visualization tools for the result and extraction function for the inquiry result must be provided. Analysis tools such as R, Python, and Zeppelin were used for data analysis, and various algorithms such as statistics, machine learning, and deep learning were used according to the data analysis environment. Through AI technology-based data analysis, Hanwoo big data platform provide data which can be practical help with Hanwoo livestock farms by developing system model. All models are predictive models created based on data provided by farms, companies, and related organization. The types of models include a breeding prediction model such as estrus, fertilization and delivery, a carcass grade prediction model related to beef quality, and a Hanwoo disease prediction model.

Establishment of an integrated log management system

A log management system that enables automated real-time monitoring is needed to safely collect, store, and analyze all log data generated by Hanwoo farms. To establish an integrated log management system, big data processing technology, real-time data processing environment, and storage suitable for the characteristics of the collected data must be provided. Accordingly, individual characteristic data such as body temperature, activity, and livestock data, such as humidity and CO₂ concentration, can be collected through a wireless network.

Establishment of an internet network connection system for data management

Unless the ICT equipment is directly developed and installed in farms, the most realistic method to receive livestock data from Hanwoo farmers is through servers from ICT equipment companies. To achieve this, consultation with an ICT equipment company must be performed and an agent that minimizes the overload on the server must be installed. Further, applied transport layer security (TLS)-based encryption must be applied between the agent and big data platform communication channel or a communication channel that applies its own certificate must be used. Internet networks connected with related organizations have additional costs, but the safest method, virtual private network (VPN), should first be considered. If VPN application is difficult, a communication channel using TLS, or a certificate method should be established.

Establishment and management plan for the regional Hanwoo digital big data center

When data can be transmitted from Hanwoo farms and ICT equipment companies using ICT equipment to select agent or agentless methods, the servers of Hanwoo farms and ICT equipment companies can be connected to the servers of the regional Hanwoo big data center. Conversely, if it is impossible to transmit data from Hanwoo farms and ICT equipment companies using ICT equipment, through ICT nodes

and ICT controllers, which are equipment and programs that help Hanwoo farms and ICT equipment companies connect servers Hanwoo data can be provided via connection with the regional Hanwoo big data center server. Accordingly, the regional Hanwoo big data center receives management, reproduction, disease, and breeding data from data producers, predicts the estrous and optimal time of fertilization in real-time to effectively manage livestock on behalf of small Hanwoo farms, and actively responds to environmental, market, and disease situations. The regional Hanwoo big data center can identify the correlation between data and help solve various livestock problems by overseeing the management data, breeding data, production data, individual data, etc. In addition, the regional Hanwoo big data center can provide auction price, shipment count, disease data, breeding data, and reproduction data to farms by subject, and data based on laws on the provision and use of public data to ICT equipment companies and related organizations (Fig 5) (Ham YH et al., 2018).

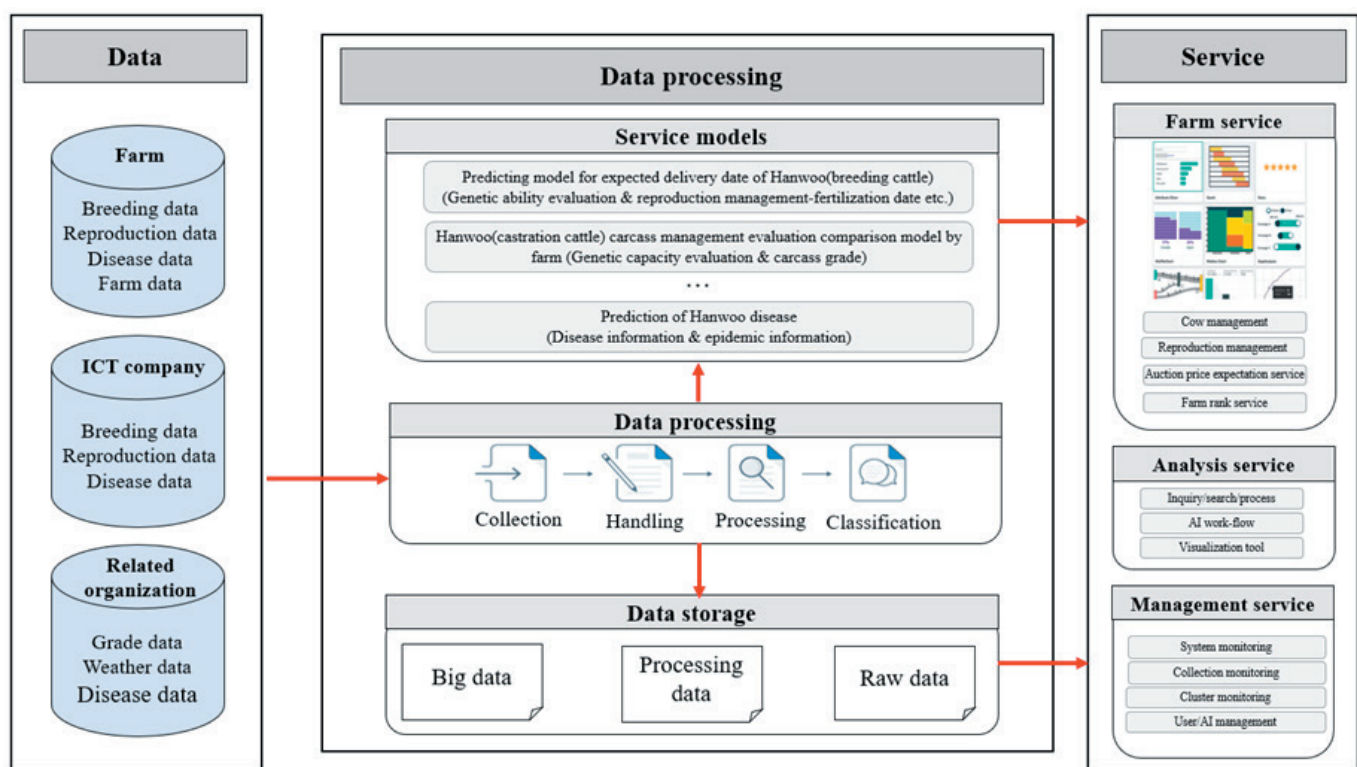


Figure 5. Flow chart of the big data-based application model

Conclusion

The establishment of regional Hanwoo big data center and system using ICT equipment is expected to have various positive effects. In particular, Hanwoo farms are expected to save on labor force needs owing to the use of ICT equipment facilities. Additional labor cost savings are also expected through the utilization of various notification services offered by regional Hanwoo big data center. The regional Hanwoo big data center is also expected to markedly contribute to the increase in the income of farms by reducing calf losses through the provision of support for cattle calving management based on estrus and dystocia prediction. According to a report by the Ministry of Science and ICT, the establishment of a big data center is expected to lead to 1.6 employees being hired per year, and the total number of employees will reach approximately 160, depending on the R&D contribution over the next five years (Song JH et al., 2018). In addition, through pedigree management using IoT technology, positive effects can be expected in other industries, such as the production of high-quality food and the development of livestock management technology using AI (Samjung KPMG Economic Research Institute, Korea, 2016). Although the modern livestock industry is larger than that of the past, many Hanwoo farms cannot focus only on farm management. Hanwoo farms are engaged in farming and pomiculture in addition to livestock industry. As a result, it is easy to miss factors that can be directly related to the livelihood of farms, such as weak estrus and respiratory diseases, which cannot be determined in real-time. The regional Hanwoo big data center can preserve farm profits by integrating disease, production, and breeding management systems to minimize the loss of small Hanwoo farms. In addition, by datafication of pedigree data, including high-quality cattle and elite cattle lists at the regional Hanwoo big data center, the loss of the Hanwoo industry can be prevented by saving high-quality cattle and elite cattle that are unnecessarily slaughtered because they are deemed low grade.

Acknowledgement

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through Advanced Production Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (318103-02).

References

- Bae HH, Son JY, Shin JS, Seo DK. 2018. Livestock technology development-Economic feasibility analysis method and case. Jeollabuk-do. Korea.
- Cheon DW, Seo DK, Son JY, Park YD. 2016. Recent livestock industry status and prospect. National Institute of Animal Science. Jeollabuk-do. Korea.
- Ham YH, Jeong JH, Song JJ, Hwang JA, Lee EJ, Heo YJ, Song JH, Jo SI, Jeon YE, Joo SK, Lee CH, Shin HC, Lee WS, Yoon HK, Lee HD, Han GS, Kim KH, Lee JS, Lee MH, No SY, Park SM, Park JH, Jeong HY, Kim SB, Lee HJ, Bok JD, Lee EB, Son YJ, Kim MA. 2018. The study on the livestock ICT standard, utilization of Big data and the technology for health monitoring of dairy cattle. Rural Development Administration. Jeollabuk-do. Korea.
- Huang J, Guo P, Xie Q, Meng X. 2015. Cloud Services Platform based on Big Data Analytics and its Application in Livestock Management and Marketing. In *Proceeding of Information Science and Cloud Computing*. 63.
- Jang YJ, Kim TW. 2019. Smart Farm Expansion and Distribution Project Status and Tasks. National Assembly Research Service. Seoul. Korea.
- Jeong YC, Han EY. 2014. Research on strategies to promote the big data industry. Korea Information Society Development Institute. Chungcheongbuk-do Korea.
- Kang MA, Kang SK, Im YS, Lee HK. 2016. U-IT-based smart dairy integrated management system development. Ministry of Agriculture, Food and Rural Affairs. Sejong. Korean.

- Kim HJ, Oh SE, Ahn SH, Choi BG. 2017. Real-time Monitoring Method of Cattles Temperature for FMD Prevention and Its Case Studies. Korean Institute of Information Technology. 15(5):141-150.
- Kim JH, Kim MC, Hwang JY, Han KN, Jeon MJ, Jeong MS. 2014. A Research on the Status of the ICT Industry Convergence and Industry Ripple Effect. Ministry of Science, ICT and Future Planning. Sejong. Korea.
- Kim JK, Oh MG, Choi DO. 2016. Design of Calf Disease Consulting System Using Big Data. The Korean Entertainment Industry Association. 48-52.
- Kim JK, Oh MG, Choi DO. 2017. Design of Big Data PlatForm for Calfscour. The Korean Entertainment Industry Association. 109-112.
- Kim JS. 2012. Consideration of big data utilization and related technologies. The Korea Contents Association. 10(1):34-40.
- Kim YJ, Seo DS, Park JY, Park YK. 2016. Research for Smart farm operation status analysis and development direction. Ministry of Agriculture, Food and Rural Affairs. Sejong. Korea.
- Korea-China Science & Technology Cooperation Center. 2018. China's Big Data Support Policy and Trends. Korea-China Science & Technology Cooperation Center. Beijing. China.
- Korea Internet & Security Agency, 2014. Analysis of Japan's Big Data Policy Implementation Status and Implications in Korea I . Global Information & Communications Technology (ICT) Broadcasting Weekly Issue, Korea Internet & Security Agency.
- KOSTAT (Statistics Korea), Korea institute for animal products quality evaluation. 2020a. Livestock trend survey. Number of farms and number of animals by province/breeding scale of korea cattle. https://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ZTITLE&parmTabId=M_01_01&outLink=Y&entrType=#content-group on 1 July 2021.
- KOSTAT (Statistics Korea), Korea institute for animal products quality evaluation. 2020b. Livestock trend survey. Number of farms and number of animals by species and province. https://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ZTITLE&parmTabId=M_01_01&outLink=Y&entrType= on 1 July 2021.
- Ministry of the Interior and Safety. 2019. Public data management guidelines. Ministry of the interior and safety notice Act No. 2019-71.
- National IT Industry Promotion Agency. 2014. The UK government's strategy for intension big data capabilities. National IT Industry Promotion Agency. Chungcheongbuk-do. Korea.
- Open Government Data Portal. 2020. Government data map. <https://www.data.go.kr/tcs/opd/ndm/view.do> on 1 July 2021.
- Park SH, Jo HS, Jang JH, Ban HJ, Park SY, 2016. Research for data experts training plan. Korea Data Agency. Seoul. Korea.
- Personal information protection commission. 2020. Personal information protection act. Personal information protection commission notice Act No. 16930.
- Pomar J, Lopez V, Pomar C. 2011. Agent-based simulation framework for virtual prototyping of advanced livestock precision feeding systems. Computers and Electronics in Agriculture. 78:88-97.
- Samjung KPMG Economic Research Institute. 2016. Smart farm industry analysis and success case. Samjung KPMG Economic Research Institute. Seoul. Korea.
- Seo DH, Hwang WS, Kim SH, Kim SM, Oh IH. 2015. Analysis of economic effects of ICT convergence. Korea Institute for Industrial Economics & Trade. Sejong. Korean.
- Seong KI, Han MH, Kim BH, Kim HG, Park KH. 2015. Analyzing and countermeasure for smart livestock farming based on ICT. Ministry of Science, ICT and Future Planning. Sejong. Korea.
- Son CM, Nam SH, Na KD, Son CH. 2019. Agricultural big data center construction plan. Gyeongbuk Provincial Government. Gyeongsangbuk-do. Korea.
- Song JH, Song MH, Park HB, Jang SA. 2018. A study on realities of conditions for the activation of data center industry ecosystem. Ministry of Science and ICT. Sejong. Korea.