Research on Big Data Model of Urban and Rural Cold Chain Storage and Distribution in Jilin Province Based on Low Carbon Perspective

Lan Gao, Xiaoming Zhang* and Yang Yang

School of Economics and Trade, JiLin Engineering Normal University, Changchun 130000, China

*Corresponding author e-mail: yaoyue1973@sina.com, 272173689@qq.com, 1026349353@qq.com

Abstract. Logistics involves production, warehousing, distribution, sales, etc., all major links have become the industrial supply chain, resulting in a large number, a wide variety and complex big data. Under the concept of low carbon and environmental protection, the urban and rural cold chain distribution in Jilin Province should also respond to the green development policy, find laws from warehousing and distribution big data, and find a path of resource conservation and cost control. Based on this, the paper relies on Hadoop to design a big data processing platform for cold chain logistics distribution and warehousing in Jilin Province, and put data blocks into data nodes to achieve cold chain logistics data analysis and improve data processing efficiency.

1. Introduction
With the continuous improvement of big data technology, it has created a good opportunity for the development of China's logistics industry. At the same time, with the rapid development of science and technology, refrigeration technology continues to advance, based on refrigeration technology, refrigeration chain technology as a means of cold chain logistics is more and more people's attention. Under the policy of continuously promoting green, low-carbon and sustainable development in the country, cold chain logistics should also need to put energy conservation and emission reduction and green development on the agenda. In recent years, the output of agricultural products in Jilin Province is generally lower than that in other provinces in the country. However, in food crops and forest products, the output of agricultural products such as rice, ginseng and pigs rank among the top in the country. The first GDP of Jilin Province increased from 72.622 billion yuan in 2011 to 164.061 billion yuan in 2018, and the demand for logistics is growing. There are various kinds of rural commodities, including agricultural products such as agriculture, forestry, animal husbandry and fishery, agricultural production materials such as pesticides, fertilizers and agricultural machinery, farmers' necessities and rural recyclables. The diversity of rural commodities determines the complexity of rural logistics. The general life must be distributed through ordinary logistics. Frozen products such as frozen products and vegetables are difficult to store and require cold chain logistics for transportation, such as agricultural machinery. Goods, but also the support of special logistics. Therefore, rural logistics presents a diversified development trend and forms a complex logistics system.
With the development of Internet technology and e-commerce, especially the development of cloud computing and Internet of Things technologies, human, sensors, PCs, servers, mobile phones and other machines generate more and more data in daily production and life. How to deal with these big data are facing huge challenges. At present, there have been many researches and designs of big data processing frameworks at home and abroad. The main big data processing frameworks include Google's Google MapReduce [1] and Apache Hadoop's Hadoop MapReduce [2]. Both of these frameworks have their own Distributed file systems, GFS [3] and HDFS [4]. In view of these demands, this paper designs an effective big data processing strategy ECL Hadoop based on Hadoop for the data-intensive application of urban and rural cold chain storage and distribution in Jilin Province.

2. A Summary of Cold Chain Logistics, Big Data Theory and Integration Theory
Cold Chain Logistics refers to a systematic project in which the frozen and frozen foods are always in the prescribed low temperature environment during production, storage, transportation, sales and consumption to ensure food quality and reduce food loss; Big data (big data) refers to a collection of data that cannot be captured, managed, and processed by conventional software tools within a certain time frame. It is a massive amount of decision-making, insight, and process optimization capabilities that require new processing models. High growth rate and diversified information assets; and integration refers to the process of gradually combining a number of originally independent sovereign entities into a single entity in some way.

![Fig.1 Big data running through the cold chain logistics system](image)

The integration theory has important guiding significance for the development of cold chain logistics. The integrated cold chain logistics model is the core cold chain logistics enterprises relying on their own advantages, relying on the market or production base to establish their own supply chain, forming agricultural products production, processing, storage and transportation, distribution and providing market information and other services. In this model, through the operation of cold chain logistics enterprises, consumers are connected with production or processing enterprises, the market will feedback the needs of consumers, and enterprises will formulate corresponding strategies through accurate market demand information to satisfy consumers. At the same time, to ensure the maximization of their own interests.
3. Design of ECL Hadoop System for Urban and Rural Cold Chain Big Data Processing Strategy in Jilin Province

3.1. ECL Hadoop system design
Urban and rural cold chain logistics big data comes from the relevant information published by urban and rural cold chain logistics service customers on various social network websites. These big data will be divided into 64MB data blocks according to Hadoop's mechanism. A time stamp will be added for all data blocks. In addition, we will add application correlation tags for each data block. These application association tags are from previous analysis requirements and are related to urban and rural cold chain logistics services. That is to say, if the analysis requirement tells us that data set A needs to be connected with data set B, we can think that some data blocks of data set A are related to some data blocks in data set B. Therefore, add an associated tag to these data blocks. After completing the semantic calculation of all the data blocks, the urban and town cold chain logistics big data placement routing table can be obtained. Through it, the placement of big data in urban and rural cold chain logistics can be completed, and those closely related data blocks will be stored in the same node. Because the shuffling phase is avoided in the processing of MapReduce, the efficiency of the calculation can be improved, especially the associated query calculation [5]. As shown below.

![Co Hadoop big data processing strategy](image)

3.2. Correlation analysis of big data in urban and rural cold chain logistics
In order to better illustrate the big data processing strategy of urban and rural cold chain logistics, the definition of big data correlation between urban and rural cold chain logistics and the calculation of big data correlation between urban and rural cold chain logistics are particularly important. The following is a brief analysis:

3.2.1. Big data correlation between urban and rural cold chain logistics. In the calculation of urban and rural cold chain logistics big data, those sets of data sets with computational dependencies are called related data sets. For example, in urban and rural cold chain logistics applications, it is assumed that there are very frequent calculation relationships as follows: calculating the amount of express delivery traffic for logistics through China Air China. As can be seen from this demand, the "order form" and the "logistics business transport company table" are two tables with a Join calculation.
relationship. Therefore, it can be seen that the “order form” and “logistics transportation company table” are two 3.2.2 urban and rural cold chain logistics big data correlation calculation.

The calculation of the correlation of big data in urban and rural cold chain logistics is mainly realized by analyzing and calculating the business needs. Suppose the big data analysis needs of urban and rural cold chain logistics companies and the table distribution they need are as follows. It can thus be seen that data set A is associated with data set C and data set D, while data set B is associated with data set E.

1 Business needs 1: {Table (A), Table (C)};
2 Business Requirements 2: {Table(A), Table(D)};
3 Business needs 3: {Table (B), Table (E)}.

3.3. System Module

3.3.1. Logistics Center Operation Module. Through the module, the cold chain logistics product information data collection and the monitoring environment parameter setting are completed; the information data collection and monitoring system configuration is completed; the current monitoring information of each monitoring point is displayed in real time, and the trend diagram of important parameters is drawn. Through this module, the operator can also initialize the monitoring hardware parameters in the system, and can choose to close unused monitoring points and monitor the running status of each monitoring point monitoring device.

3.3.2. Agent module. The module is the core of the entire cold chain logistics system. It includes components such as management agent, decision agent, communication agent, monitoring agent, and execution agent. The module is responsible for real-time monitoring and intelligent control of logistics products, and drawing waveforms for monitoring environmental changes. The management agent is responsible for the supervision, control and management of the entire system, and coordinates and schedules the work between other agents. Therefore, it is represented by a hybrid logical agent.

3.3.3. Fault Alarm Module. When a certain indicator exceeds the normal value during the running of the system, an automatic alarm is generated to generate a log file; when the monitoring device fails, a fault signal is issued, or the controller is out of control, the monitoring system issues an alarm display to automatically store the alarm data.

3.3.4. Big Data Analysis Module. Based on the big data characteristics of food safety, Hadoop and MapReduce system architecture, based on NoSQL data management technology, Hive is used for ETL data processing, including data acquisition, data processing, data analysis, data presentation and application modules.
Fig.3 System big data monitoring and acquisition system design

(1) Data acquisition: not only includes routine monitoring of foodborne disease surveillance data, food pollutant monitoring data, cold chain temperature and humidity and other structured data, but also remote sensing data related to food safety environmental pollution, food safety public opinion monitoring data, etc. Unstructured data; breakthroughs in the “fences” and “islands” of data from agriculture, quality inspection, industry and commerce, and health, to achieve “liquidity” and “availability” of data.

(2) Data processing: Using the respective advantages of RDBMS and NoSQL, using the form of mixed storage of structured and unstructured data, mutual compensation for defects, achieving flexible and efficient purposes.

(3) Data analysis: Using spatial statistical analysis methods to explore the correlation between various impact factors and various food safety risks, and establish a multi-factor influence model; using complex network technologies, between multiple impact factors Intrinsic associations and interactions are explored and modeled to construct a complex network model of multiple impact factors. Based on a unified space-time framework, based on the complex correlation between impact factors and food safety risks, a multivariate impact factor is established. Spatio-temporal reasoning models, etc.

(4) Data presentation and application: Application GIS transmits the food safety event site information back to the data center display, and displays the real-time data change situation and related resource distribution in the basic space geographic image, through spatial distribution analysis and time. Historical curve analysis, buffer analysis, shortest path analysis, etc., time-space distribution and development trend of events, decision-making, early warning and forecasting.

3.4. Big Data Acquisition Algorithm
The coordination and cooperation between the modules are expressed by algorithms. It includes both the behaviour of a single module and the behaviour of cooperation by multiple modules. It uses the transfer function to realize the information exchange between each module. The problem-solving process and the data acquisition process in the multi-module system adopt a cooperative algorithm. There is a task that can be divided into n subtasks, which are executed by n tasks respectively. The execution time of the task is limited to a certain time window area, and all tasks completed outside this
area have to pay extra cost [6]. A task execution Agent \( j \) contains three important indicators \((A_j, B_j, P_j)\), where \( A_j \) is the cost required for task \( j \) to complete before the time window, \( B_j \) is the cost required for task \( j \) to complete after the time window, and \( P_j \) is executed. The time required for task \( j \). The time window for completing the total task is \([t-a, t+a]\), where \( 2a \) is the window size, \( t \) is the best completion task time, and \( t \) value is to be determined. Solving the appropriate scheduling scheme through the scheduling module minimizes the extra cost of performing the task. Its objective function is expressed as follows [5].

\[
S = \sum_{j} \left[ A_j \cdot q(t-a-t_j) + B_j \cdot q(t_j - a - t) \right]
\]

\[
\max \sum_{j \in L} S_j X_j
\]

\[
s.t \sum_{j \in L} P_j X_j \leq 2a \quad X_j \in \{0, 1\} (j \in L)
\]

Where \( t_j \) is the completion time of task \( j \). Each of these task completion times is assumed to be the start time of the common completion task time window. As a task completed in the time window in this scheduling, it selects the minimum value of the objective function value among the \( n \) subproblems as an optimal value of \( S \), and determines the optimal solution of \( S \).

4. Conclusion
Building a public information platform for urban and rural cold chain logistics based on cloud computing and big data, which can meet the growing demand for urban and rural cold chain logistics, realize the effective connection between urban and rural two-way logistics, and form a two-way circulation urban and rural cold chain logistics development model, thereby reducing logistics resources. Waste, improve logistics and distribution efficiency. In addition, the urban and rural cold chain logistics public information platform uses advanced management concepts and information technology to innovate the urban and rural cold chain logistics model, which can reduce urban and rural cold chain logistics links, reduce agricultural product circulation loss, protect food safety for urban residents, and promote farmers' income. Important role.

Acknowledgments
This work was financially funded by science research project of education department of jilin province, Project number: JJKH20190775SK, JJKH20190785SK; Supported by Program for Innovative Research Team of JiLin Engineering Normal University.

References
[1] Guo Shuangying, Chen Mingjing, Shen Diyi. Application of Big Data in Cold Chain Logistics. Modernization of Shopping Mall, Vol. 9 (2014) No.26, p. 40-42.
[2] Wu Wenshu. Analysis of the Influence of Big Data Environment on Cold Chain Logistics Distribution. Journal of Logistics Science, Vol. 10 (2016) No.39, p. 77-78.
[3] Xue Wei. Analysis of the Application of Big Data Technology in Cold Chain Logistics. Logistics Engineering and Management, Vol. 6 (2018) No.40, p. 104-105.
[4] Xie Weihang. Analysis of Real-time Monitoring and Optimization under Cold Data Logistics Big Data. Jiangxi Building Materials, Vol. 1 (2017) No.23, p. 286-289.
[5] Wang Yu, Zhao Caixiu, Xu Zitian. Research on the Integration Model of Agricultural Products Cold Chain Logistics in the Age of Big Data. China Business Review, Vol. 5 (2017) No.35, p. 23-24.
[6] Gu Li. Development Strategy of Cold Chain Logistics of Agricultural Products Based on Low Carbon Perspective. Economic Research Guide, Vol. 5 (2015) No.21, p. 28-30.