

RESEARCH

Open Access



# Intelligent algorithms for cold chain logistics distribution optimization based on big data cloud computing analysis

Yi-hua Chen

## Abstract

In recent years, the rapid development of fresh food e-commerce in China has brought about more development opportunities for the cold chain logistics industry but has also presented new challenges. With the development of cloud computing and big data technology, it is increasingly important to study the application of big data and cloud computing technology in cold chain logistics. The purpose of this paper is to study the intelligent algorithm of cold chain logistics distribution optimization based on big data cloud computing analysis. Based on the constituent elements of the cold chain distribution problem and using cloud computing technology to obtain real-time traffic information in the transportation system through a unified access interface, this article analyses the distribution time and cost of refrigerated vehicles, thereby establishing a cold chain distribution vehicle path optimization model. By analysing the parallel programming mode of cloud computing, the parallel design and analysis of a coarse-grained genetic algorithm are used to solve the simulation model of the established optimization model. The experimental results show that the method of optimizing cold chain logistics vehicle routing using cloud computing is effective. When comparing 1, 2, 4, and 8 processors, the execution times are 19.89, 14.52, 8.12, and 6.41, respectively. It can be seen that the more processors there are, the shorter the calculation time.

**Keywords:** Big data, Cloud computing, Cold chain logistics, Distribution optimization

## Introduction

With the improvement of people's living standards, the structure and nutritional value of foods have also received increasing attention. The residents of our country are also slowly changing their diet structure. The diversified and varied consumption patterns of fruits, vegetables, quick-frozen products, and eggs and poultry have changed from the previous single consumption structure for food and clothing to a diversified food structure for the pursuit of balanced nutrition and health. The proportion of fresh food in people's daily diet is also increasing. Fresh food is difficult to keep fresh at room temperature, so it needs to be distributed from the

origin to the consumer terminal through cold chain logistics [1]. This often requires cold chain logistics technical support in various tasks, such as warehousing, transportation, and distribution. The frequent food safety issues in recent years have caused the country and residents to attach great importance to food safety [2]. In people's minds, dietary health has become a key issue in life, especially concerning the quality and safety of raw, cold and fresh food. The degree of attention to this topic has gradually increased. If you want to ensure the quality of fresh food, then you must ensure its safety and smoothness in the circulation process, that is, to ensure the traceability of fresh food circulation information, including fast speed, quality and quantity, to further meet the needs of consumers in consuming fresh food. In recent years, with the continuous deepening of Internet

Correspondence: [cy89aa@163.com](mailto:cy89aa@163.com)  
School of software, South China Normal University, Nanhai 528225,  
Guangdong, China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

technology, the degree of informatization of logistics enterprise management has continued to increase. The emergence of information technologies such as cloud computing and the Internet of Things has led to good results for enterprises in public logistics information platforms and intelligent logistics system application prospects [3]. The application of big data and cloud computing in logistics is based on the integration of logistics resources and capabilities, dynamic on-demand services, and integrated logistics service requirements [4]. Cloud computing can quickly search for real-time and dynamic information in a short period of time, can transfer the latest vehicle transportation, storage status, real-time routes and other information to logistics centres and vehicles, and obtain the calculation results in a short time, thus improving the efficiency of distribution and increasing economic benefits. This paper is aimed at the problem of the real-time route optimization of cold chain logistics distribution vehicles, uses cloud computing processing methods for analysis, establishes an application service that is consistent with the actual distribution of cold chain logistics vehicles, builds an optimization model that meets the constraints, and provides planning and vehicle scheduling for cold chain logistics to provide support for optimization. Real-time route optimization services for cold chain logistics distribution vehicles based on cloud computing are used to realize distribution resources, distribution capabilities, and distribution knowledge sharing and on-demand use, improve the utilization rate of distribution resources, meet the personalized needs of users for services, and promote energy conservation and emissions reduction to achieve green and low-carbon manufacturing.

Lv and his team believe that accurate and timely traffic flow information is critical to the successful deployment of intelligent transportation systems [5]. In the past few years, the amount of traffic data has exploded, and we have truly entered the era of big data in transportation. The existing traffic flow prediction methods mainly use shallow traffic prediction models and are still unsatisfactory for many practical applications. This situation inspired the above authors to rethink the problem of traffic flow prediction based on deep architecture models with large traffic data. They proposed a new deep learning-based traffic flow prediction method that inherently considers spatiotemporal correlation [6]. Stacked autoencoder models are used to learn general traffic flow features and are trained in a greedy layered manner. To the best knowledge of the above authors, this is the first time that a deep architecture model has been applied using an autoencoder as a building block to represent traffic flow characteristics for prediction. In addition, they found that the proposed traffic flow prediction method has superior performance [7]. Dejene and his

team see cloud computing as an emerging paradigm that provides computing resources as services over the network. Communication resources often become a bottleneck for the provision of many cloud application services. Therefore, copying data (such as a database) closer to the data consumer (such as a cloud application) is considered a promising solution, which allows for the minimization of network latency and bandwidth usage. The above authors studied data replication in cloud computing data centres. Unlike other methods available in the literature, in addition to improving the quality of service (QoS) due to reduced communication delays, they also consider the energy efficiency and bandwidth consumption of the system. The evaluation results obtained during extensive simulations help reveal the tradeoff between performance and energy efficiency and guide the design of future data replication solutions [8]. G.W. Lailossa believes that cold chain management is an important factor in ensuring the quality and safety standards of fishery products. Global trends in fishery product standard requirements (quality, safety, and traceability) are increasing and becoming a global influence. These issues have led to the development of a new paradigm for tuna cold chain system management, which must be holistic, integrated and up-to-date. The review results of previous studies indicate that cold chain management is still partially applied and has not been fully integrated into the entire process (postharvest processing, processing and packaging, refrigeration and distribution, refrigerated transportation, and fishery sales). He provides a new paradigm for providing cold chain management capabilities through a system integration approach. The output of this integrated model is expected to improve the export competitiveness of Indonesian fishery products (tuna) [9].

This article analyses the components of the cold chain distribution problem, uses cloud computing technology to obtain real-time traffic information in the transportation system in a unified access interface, and then analyses the distribution time and cost of refrigerated vehicles to establish a cold chain. The chain distribution vehicle routing optimization model is analysed in parallel with the cloud computing parallel programming model. The coarse-grained genetic algorithm is designed and analysed in parallel and used to solve the simulation of the established optimization model. From the experimental results, it can be seen that the method of optimizing cold chain logistics vehicle routing using cloud computing is effective.

## **Proposed method**

### **Cold chain logistics**

#### *Definition of cold chain logistics*

Cold chain logistics refers to perishable and perishable products such as agricultural products, meat delicatessens,

and dairy products that must remain in a low-temperature environment [10, 11].

#### **Characteristics of cold chain logistics**

The main characteristics of cold chain logistics are as follows.

- 1) Storage and transportation products are perishable and perishable

During the storage and transportation process, with the passage of time, the deterioration of storage and transportation products is accelerated. Once the product deteriorates and the customer is unwilling to purchase it, the enterprise experiences a large amount of waste. Different products have different temperature requirements for storage and transportation; some require refrigerated processing, and some require frozen processing.

- 2) Large investment in cold chain infrastructure

To ensure the quality of cold chain logistics production [12], storage and transportation products and to keep them in a low-temperature environment, the products need to be stored in cold storage, and a large amount of refrigeration equipment must be purchased to monitor the temperature in real time. The daily maintenance of cold chain equipment and the training of personnel using refrigeration equipment require significant investment. Perishable products require low temperature throughout the whole process from procurement, production, and processing to transportation and distribution, especially the connection of each link. Therefore, cold chain logistics requires more investment and management funds than room temperature logistics.

- 3) High level of cold chain logistics technology and information technology requirements

Cold chain products must be kept in a low-temperature environment due to their own particularities [13]; otherwise, the products will rot and deteriorate, which will affect sales and reduce customer satisfaction. The whole process of cold chain products from procurement, production, storage, and transportation involves many technologies, such as refrigeration technology, the selection of refrigerants, the management of refrigeration systems, and vehicle refrigeration technology. To maintain the coordination of the cold chain and the freshness of the product, it is necessary to monitor the real-time status of the cold chain throughout the process. This process requires the use of RFID, GPS, GIS, GPRS and other logistics technologies.

#### **Cold chain logistics composition**

**Freezing** Cooling is the process of manually lowering the temperature using a refrigerant. Precooling refers to the process of quickly lowering the temperature of products that need to be refrigerated and transported to the final temperature in advance. This approach can extend the shelf life of fruits and vegetables and reduce their drying during circulation and transportation. Consumption and loss ensure that consumers can buy fresh green products.

**Frozen storage** Frozen storage is a process after freezing processing. On the one hand, it is used to keep the processed products in a low-temperature environment, to keep the products fresh and to prevent decay and deterioration. On the other hand, it is used to transport the products processed in the first step to a distribution centre or to the merchant for storage, which can alleviate the imbalance in production and sales and ensure that goods are available to consumers during the off-peak season.

**Refrigerated transportation and distribution** Refrigerated transportation and distribution refer to the low-temperature transportation and distribution of products from the entire logistics link of procurement, production processing, and distribution. This process involves the selection of transportation means according to the transportation products and distance. Temperature and humidity are the key factors affecting product quality. These factors must be controlled for during transportation and distribution, so the good performance of the vehicle and the real-time monitoring of temperature and humidity on the vehicle are important. In addition to the two main reasons—temperature and humidity—road conditions, wear and tear during loading and unloading, and the conversion between different modes of transportation will all affect the quality of refrigerated transportation and distribution products, chain equipment and cold chain technology.

**Frozen sales** Frozen sales are the last link of cold chain logistics. After a series of links, the products are finally sent to retail stores or hypermarkets. They are sold in refrigerated storage and refrigerated display cabinets in these malls for a short period of time. Only when the product is sold to consumers can the whole process of this product be considered complete, and then, the producers and operators can obtain profits. After the product is sold, the market feedback information can provide producers and operators with subsequent decision-making information on how the product is produced, sold, and operated; then, a new round of operation can begin.

**Cloud computing**

Cloud computing is a resource service model that combines grid computing [14], distributed computing, parallel computing, and the Internet. Its objective is to virtualize soft and hard resources such as computing, storage, and services. Software and other services integrate computing models that integrate distributed computing resources such as scalable computing, storage, data, and applications for collaborative work [15, 16]. The cloud computing architecture consists of three parts: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The architecture of cloud computing is shown in Fig. 1.

The principle of cloud computing is to use distributed computing resources to provide users with various services. Users can obtain service resources anytime and anywhere according to their actual needs. Cloud computing uses Internet and virtualization technology to connect a large number of PCs or servers together; they form large clusters, virtualize resource pools, allocate resources according to user needs, and obtain higher resource utilization rates at lower costs [17, 18]. Cloud computing, as a new application model, breaks down the resource constraints under the traditional model, stores data, applications, and services in the cloud, provides powerful computing and processing capabilities, and enables the self-adaptation of business systems [19]. As long as users can access the network, they can obtain the required service resources (computing, storage, and software).

The resource layer provides infrastructure cloud computing services [20], forming a variety of physical devices,

such as servers, networks, and storage devices, through virtualization technology to form dynamic virtual resources and provide them to users as network services.

The platform layer provides users with the encapsulation of resource layer services, enabling users to build their own applications using more advanced services. This layer is mainly designed for developers and includes parallel programming and development environments, middleware services, and distributed data management.

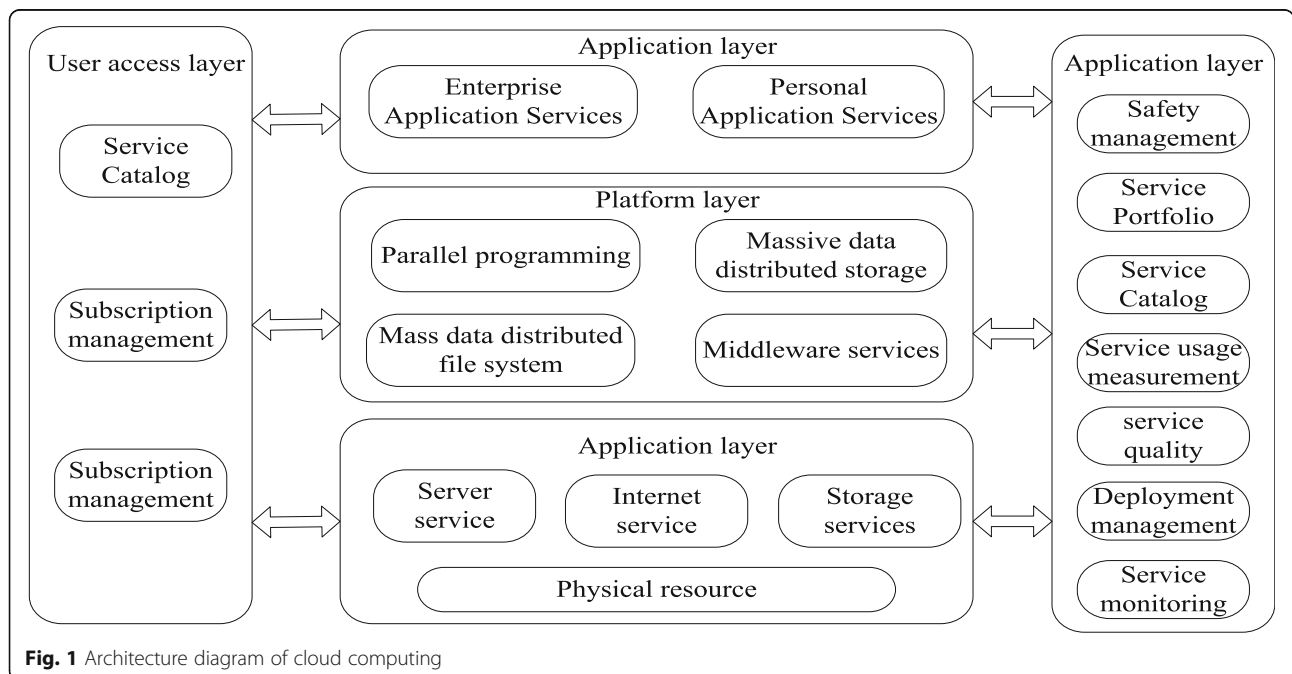
The application layer provides users with software services and interactive interfaces. Users can lease corresponding software services and customize these services according to their needs. The cloud provides the corresponding infrastructure, services, software and hardware resources. There is no need to pay for expensive hardware equipment, no need to bear frequent maintenance and upgrade costs, and no need to build their own server centre, data centre, or large computer room, thereby greatly reducing the cost of information system construction [21, 22].

The user access layer provides various support services for users to utilize the various services of cloud computing and provides various levels of access interfaces [23].

The management layer manages the cloud computing services provided at all levels.

**Big data**

The scope of the exploration of big data technology is wider than that of traditional data management technology. It includes not only structured data within a known range but also the degree of data correlation. It is specifically manifested in three points: the first is to store a



**Fig. 1** Architecture diagram of cloud computing

large amount of structured and unstructured data; the second is to process a large amount of data in real time; and the third is to build an algorithm model and continuously optimize it based on real-time data [24, 25]. A typical big data technology architecture is shown in Fig. 2.

**Big data processing**

Mining the value contained in data is the core driving force behind the development of big data technology. Big data platforms face the challenge of processing massive amounts of data in a short period of time. When the system is busy, there will be more than 100,000 or even millions of online data processing requests per second, and thus, it is necessary to use big data real-time processing technology. Big data require different processing forms for three different types of data: the batch processing of static data, real-time processing of online data, and comprehensive processing of image data.

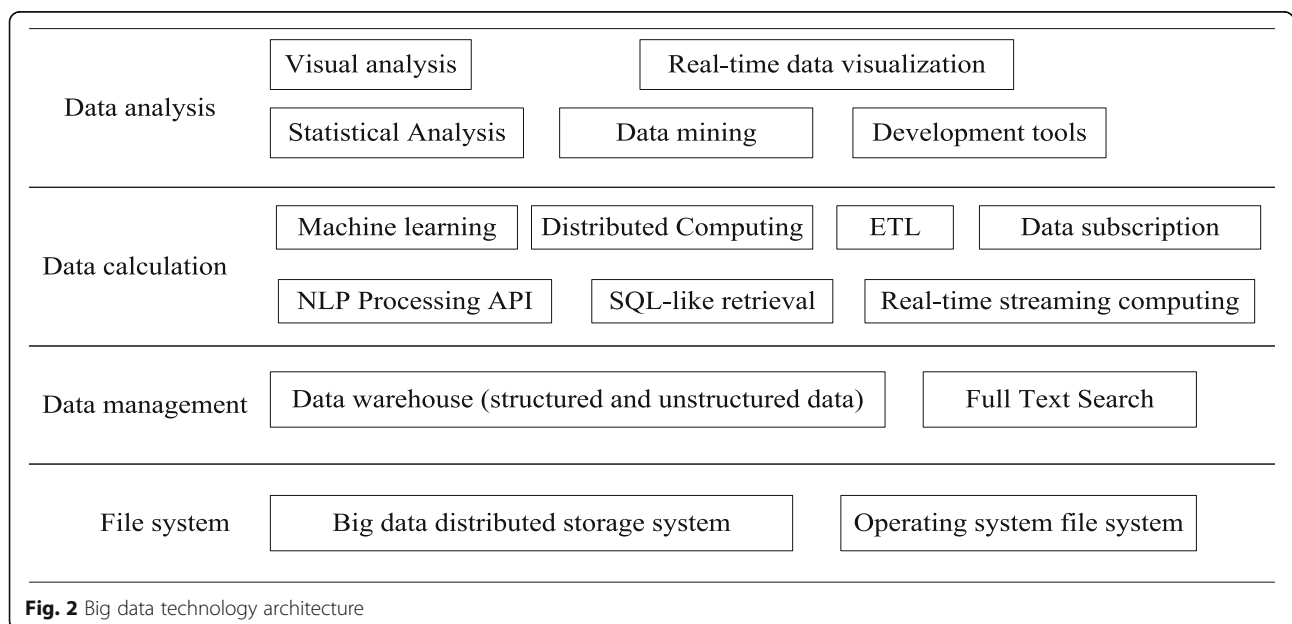
Static data are stored on storage media in a static form, and their accumulation process reflects the continuous precipitation of enterprise data assets. Due to the large volume of static data, ranging from the TB level to the PB level, the data accuracy is high, but the value density is low. It also takes considerable time to process static data and basically does not provide user-system interaction operations. The batch processing of static data is suitable for behaviour analysis in social networks, product recommendations in e-commerce, and cost and efficiency analysis in public service fields.

Technologies suitable for the batch processing of static data include the MapReduce computing framework, which was first proposed and promoted by Google. It

has great advantages in processing large data in a highly parallel and scalable manner. The programming interface is simple. You can use cheap X86 servers to build a large-scale big data framework with strong input and output (I/O) capabilities, which is easy to understand and use. MapReduce uses the working principle of decomposition first (Map) and then merge processing (Reduce). It can divide large data files and distribute them to multiple computing nodes for parallel processing. After processing, they are summarized. This data processing technology greatly improves the processing speed and has great scalability and high availability.

**Big data analysis**

Big data analysis is used to conduct in-depth observations of data to discover the relationships, patterns and trends that are valuable for decision-making and then to use discovery to establish decision-making models and provide predictive support methods and processes. In the process of big data analysis, appropriate statistical analysis methods and tools are used to extract the most valuable information from the collected data and play a role in the status quo analysis, cause analysis, and quantitative predictive analysis. Before performing data analysis tasks, clear analysis goals and assumptions need to be set, and then, whether the assumptions are correct needs to be verified through comparative analysis, group analysis, cross analysis, regression analysis and other methods; next, the analysis results need to be interpreted, and the corresponding data analysis conclusions can be drawn.



**Fig. 2** Big data technology architecture

**Optimization model of cold chain logistics distribution vehicle routing based on big data cloud computing**  
**Distribution time analysis**

By using the real-time road condition of vehicle travel, the driving speed of relevant sections of distribution vehicles is obtained, and the travel time of distribution vehicles is calculated so that the vehicles can choose the route with the shortest time possible in the process of driving. In the construction of the model, real-time road condition information is acquired. Through the unified access interface of the established service architecture, the real-time traffic information of the road segment in the city where the vehicles are distributed is obtained in the traffic information cloud, and the average driving speed of the vehicles in this road segment,  $V_1$ , lane width  $L_{Ri}$  and the tonnage of permitted vehicles are obtained. On the basis of the value of  $V_i$ , the road traffic congestion degree, when  $V_i = 0$ , the current state of congestion high; only when  $L_{Ri}$  lane width is greater than vehicle width  $L$  and allows for traffic tonnage  $Q_{Ri}$ , meaning the gross tonnage of  $Q$  is greater than that of the car, can I select the path. Therefore, the time needed for the vehicle to complete the distribution of total travel time  $T_s$  is calculated by Formula (1):

$$T_s = \sum_{i=1}^N t_{i(i-1)} \tag{1}$$

where  $t_{i(i-1)}$  is the travel time of the refrigerated truck from customer  $i-1$  to  $i$ .

**Distribution vehicle cost analysis**

According to the characteristics of cold chain logistics, this paper deeply discusses the cost of goods damage caused by low-temperature products due to their corrosivity, the energy cost consumed by refrigeration equipment during vehicle distribution, the penalty cost beyond the customer’s time window, the inherent cost of distribution vehicles and the transportation cost on each road section.

**Fixed costs** There are fixed costs to be borne by transport vehicles, namely, the wages of drivers and escorts and the cost of vehicle wear and tear, and the total fixed cost  $C_1 = f$  is constant.

**Transportation cost** In the distribution process of low-temperature products, there is a large difference in vehicle driving speed in different time periods, so the fuel consumption of vehicles also has a large difference. Especially during rush hour when traffic congestion occurs, vehicle fuel consumption will be significantly higher. The transportation costs of the vehicle includes the fuel consumption and repair and maintenance costs of the

vehicle, which are dynamic and proportional to the mileage of the vehicle. For transport cost expressed by  $C_2$ , Formula (2) can be obtained.

$$C_2 = \sum_{i=0}^n \sum_{j=0}^n c_{ij} x_{ij} \tag{2}$$

$c_{ij}$  is the transport cost of refrigerated vehicles on section  $(v_i, v_j)$ , and  $c_{ij} = c_{ji}$ , where  $x_{ij}$  is represented by 0,1.  $x_{ij} = 1$  means refrigerated vehicles have passed section  $(v_i, v_j)$ ; otherwise,  $x_{ij} = 0$ .

**Damage cost of cold storage** The damage cost of the cold collection is analysed in the following three cases. First, cargo damage is caused by the length of the transportation time during transportation. Second, when serving customers, the air flow caused by the opening of the compartment door, through which the cold air in the compartment and the outside air flow alternately, causes the temperature in the compartment to rise and damages the cold storage. Third, during transportation, the vehicle can be bumped due to the quality of the road, and the product may be damaged. The cost of cold storage damage is expressed by  $C_3$ , which can be calculated by Formula (3):

$$C_3 = r \sum_{j=0}^n \lambda_j (\alpha_1 t_{ij} + \alpha_2 \beta_1 + \alpha_3 S_{ij}) \tag{3}$$

where  $r$  represents the unit price of the product;  $\lambda_j$  is a 0,1 variable, with  $\lambda_j = 1$  representing the refrigerated truck serving customer  $j$ , and otherwise,  $\lambda_j = 0$ ;  $\alpha_1$  is the percentage of damage during the product delivery process;  $t_{ij}$  represents the time from customer  $i$  to customer  $j$ ;  $\alpha_2$  represents the percentage of product damage during the door opening and loading/unloading process;  $\beta_j$  is the quantity of customer  $j$ ’s cargo;  $\alpha_3$  is the percentage of product damage during vehicle transportation; and  $s_{ij}$  is the mileage between customer  $i$  and customer  $j$ .

**Energy costs of refrigerated vehicles** The cost of energy is mainly the cost of consuming refrigerant. The consumption of refrigerant is mainly related to the heat transfer coefficient of the cabin, the temperature inside the cabin, the surface area inside and outside the cabin, and the outside temperature. The amount of refrigerant consumption  $G$  can be calculated by Formula (4):

$$G = a \times b \times S \times \Delta t \tag{4}$$

where  $G$  represents the consumption of refrigerant,  $a$  is a constant,  $b$  represents the heat transfer coefficient,  $S$  is the average surface area inside and outside the vehicle, and  $\Delta t$  is the temperature difference between the inside and outside of the vehicle. Then, the cooling cost of the

vehicle during driving is represented by  $C_4$ , which can be calculated by Formula (5):

$$C_4 = \sum_{i=0}^n \sum_{j=0}^n r_1 \times G \times t_{ij} \times x_{ij} \quad (5)$$

where  $r_1$  is the price of the refrigerant,  $t_{ij}$  is the time from customer  $i$  to customer  $j$ , and  $x_{ij}$  is a variable of 0,1. In addition, when the door is opened, the outside air directly convects with the air in the refrigerated car to alternate between cold and heat. Therefore, when calculating the cooling cost  $C_5$  of the door, only the cost of the refrigerant consumed by the heat exchange of the door is calculated, such as in Formula (6):

$$C_5 = \sum_{i=1}^n r_1 \times a \times S \times \Delta t \times t_i \quad (6)$$

where  $t_i$  is the time the vehicle stays and waits at customer  $i$ , and  $S$  is the area of the door at this time.

**Penalty costs beyond customer delivery time** The delivery of goods by logistics companies to customers is limited by time. Here, we use the limitation method of soft time windows for optimization; that is, the goods required by customers must arrive within a certain time range. If they fail to arrive on time, then customers will impose a fine on the delivery company. To prevent the refrigerated truck from unloading the goods earlier than the time window, we adopt the early arrival waiting principle; that is, if the time of arrival of the goods is earlier than the start time of the customer's time window, then the logistics vehicle should wait at the customer until the start of the customer's time window and then unload. Then, the penalty cost incurred by the customer because the delivery task is not within the time window range is represented by  $C_6$ , which can be calculated by Formula (7):

$$C_6 = \omega_1 \sum_{j=1}^n \max[(a_j - s_j - t_j), 0] + \omega_2 \sum_{j=1}^n \max[(s_j - b_j), 0] \quad (7)$$

where  $\omega_1$ ,  $\omega_2$  represent the loss cost caused by the refrigerated vehicle unloading the goods earlier and later than the time window, respectively.  $s_j$  is the time when the vehicle arrives at customer  $j$ , and  $t_j$  is the time when the vehicle waits at customer  $j$ .

**Other costs** In the process of the distribution of goods, under time constraints, vehicles will sometimes travel on expressways, which will generate high-speed charging costs. At the same time, some cities will levy different

traffic congestion charges when vehicles drive during peak periods. These costs are expressed as  $C_7$  and can be obtained by Formula (8).

$$C_7 = \sum_{j=1}^m g_j \quad (8)$$

### Establish an optimized path model for cold chain logistics distribution vehicles

The path decision model of vehicle transport in cold chain logistics is constructed as follows:

$$\text{Min}\{W_1B + W_2C\} \quad (9)$$

Among them,

$$B = T_s + \sum_{i=1}^N t_j \quad (10)$$

$$C = C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 \quad (11)$$

## Experiments

### Data collection

The data are on the cold chain logistics distribution products of a supply chain management company, which now services more than 560 restaurants, with the daily servicing of 300 restaurants. Due to the large number of service stores, only 12 restaurants are selected as the demand stores for the cold chain logistics distribution service studied in this paper, and the distribution centre of the company completes the distribution task of these 12 demand stores. In addition, the company's distribution centre number is 0, and the store number for each demand is 1,2,..., 12.

### Experimental design and experimental steps

#### Parallel programming mode

In the context of big data cloud computing, to allow users to quickly and easily obtain services in cloud computing and make full use of cloud computing resources, a parallel programming model is used to enable large and complex tasks in the background to be executed in parallel, hiding the parallel from developers details of distributed processing, such as virtualization, fault tolerance, and data distribution. Map-Reduce is a simplified parallel programming model that has the advantages of the simple and effective calculation of large amounts of data.

The Map-Reduce system consists of three modules. The client-side Map-Reduce API submits parallel processing jobs written by developers to the Master node; the Master node automatically decomposes user jobs into map and reduce tasks, and then, the tasks are

scheduled to worker nodes; and the work tasks of worker nodes are requested in the Master node. In addition, the input/output data of Map-Reduce can be stored in a distributed file system composed of multiple worker nodes. The cloud computing parallel execution process is shown in Fig. 3.

**Solution of optimization model by parallel genetic algorithm**

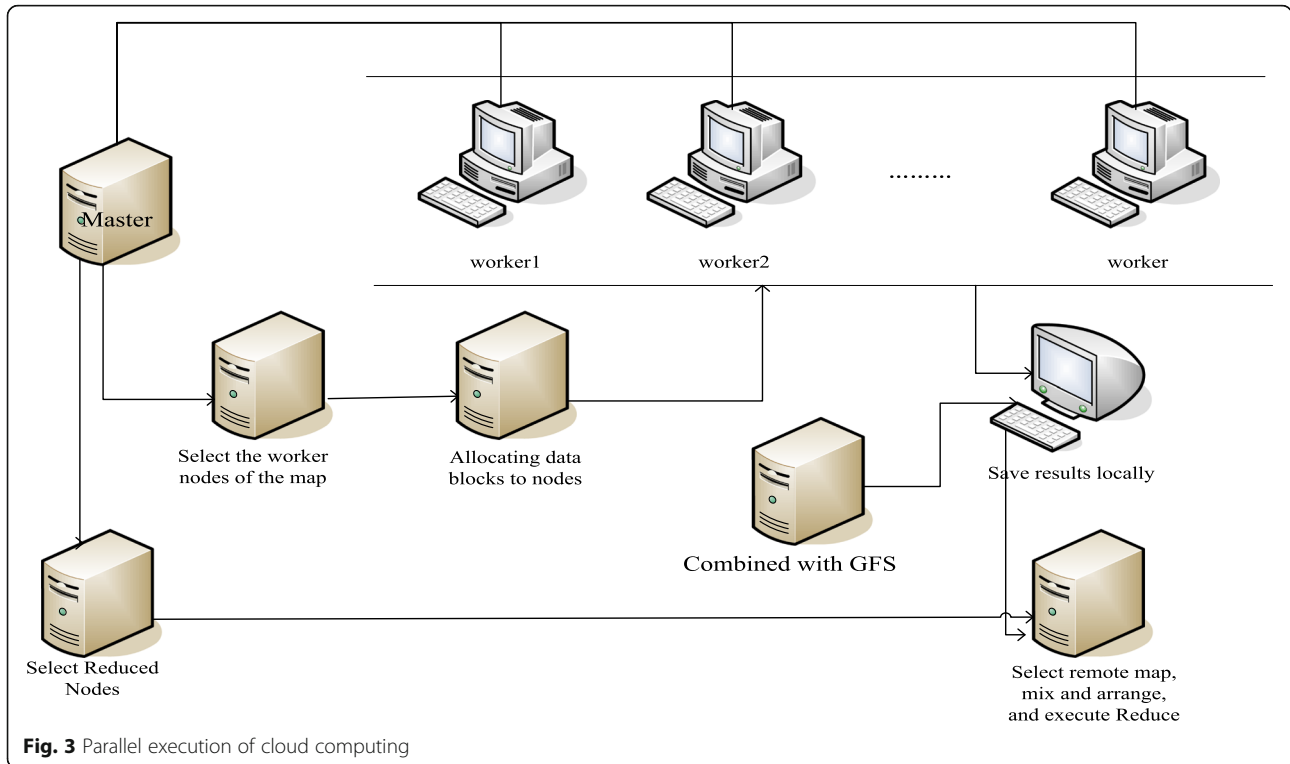
In the context of big data cloud computing, this paper uses the genetic algorithm coarse-grained parallel mode to solve the real-time optimal path model of cold chain logistics distribution vehicles. The coarse-grained genetic algorithm divides the randomly generated initial population into several subpopulations according to the number of processors. Each subpopulation executes evolution operations concurrently and independently on different processors. Each time after a certain evolutionary generation, several individuals will be exchanged to introduce excellent genes from other subgroups, enrich the diversity of each subgroup, and prevent the occurrence of immature convergence; moreover, the communication overhead of the coarse-grained model is small, and a nearly linear acceleration ratio can be obtained, which is suitable for running in communication on a low-bandwidth cluster system.

$m \times n$  random subpopulations are distributed in the feasible set, each subpopulation has  $n$  individuals, and each subpopulation is independently genetically evolved

for a certain number of generations. The sampling points are concentrated on the optimal solution through the information transfer between subpopulations. The search space of each subgroup will be smaller and smaller. In this way, each subpopulation is mapped to  $m$  processors. These processors independently perform genetic steps in parallel; that is, they independently perform selection, crossover, and mutation operations. After selection, hybridization, and mutation, a new generation of individuals is added as the current population. The fitness value of the function is calculated, and individuals with poor performance are eliminated. Because there are multiple groups and the search space shrinks at a different rate for each group, it can be guaranteed that the search space always contains the optimal solution. Each subgroup will eventually find one or more local extreme points.

The steps of genetic algorithm execution in the big data cloud computing environment are as follows:

- Step 1: Population initialization
- Step 2: Divide the initial population into  $m$  subpopulations
- Step 3: The main loop starts while  $i \neq m$  do
- Step 3.1: While  $j \neq n$  do
- Step 3.1.1: Chromosome coding
- Step 3.1.2: Calculate the fitness value of the function
- Step 3.1.3: Select
- Step 3.1.4: Cross



**Fig. 3** Parallel execution of cloud computing



**Table 1** Comparison of optimization results

Index	Cost before optimization (\$/day)	Optimized cost (\$/day)	Reduction ratio (%)
Fixed cost	1645	1136	20
Variable cost	5457	4361	16.7
Cooling cost	1620	1110	20.1
Damage cost	1367	967	15.7
Penalty cost	992	576	24
Total cost	11,581	8150	19

Step 3.1.5: Mutation

Step 3.1.6: Extract individuals to migrate from the subpopulation

End while

Step 3.2: Migration

End while

Step 4: Output results

## Discussion

### Genetic algorithm optimization process and result analysis

#### *Analysis of genetic algorithm optimization results*

The model constructed in this study is based on the minimum total cost as the optimization goal. After 200 iterations, the approximate optimal solution of the model is 8150. The fixed cost, variable cost, refrigeration cost, damage cost, penalty cost and total cost are compared. An index is used to compare the results before and after optimization. Table 1 and Fig. 4 show the comparison of the optimization results. We can see that compared with the previous total cost of 11,581, a reduction of 3431 yuan is achieved; that is, the total cost of the vehicle is reduced by 19%. The comparison of specific optimization results is shown in Table 1 and Fig. 4.

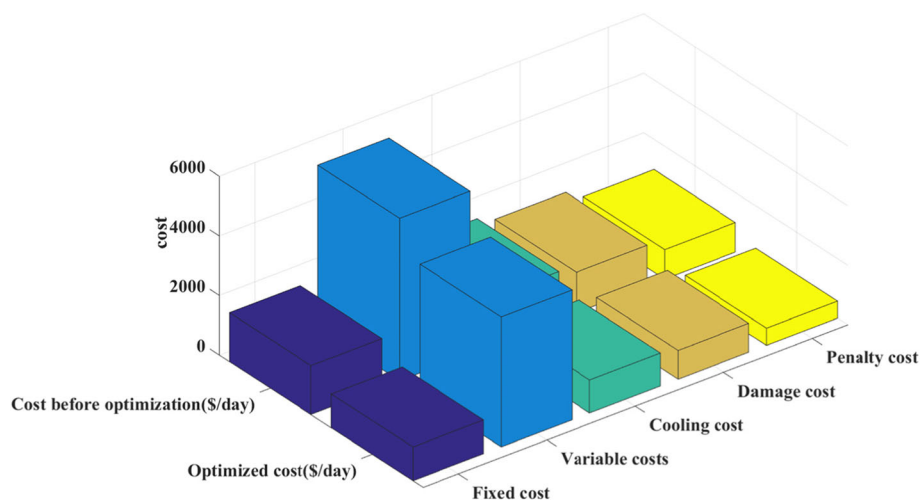
### *Comparative analysis of the number of iterations and average cost*

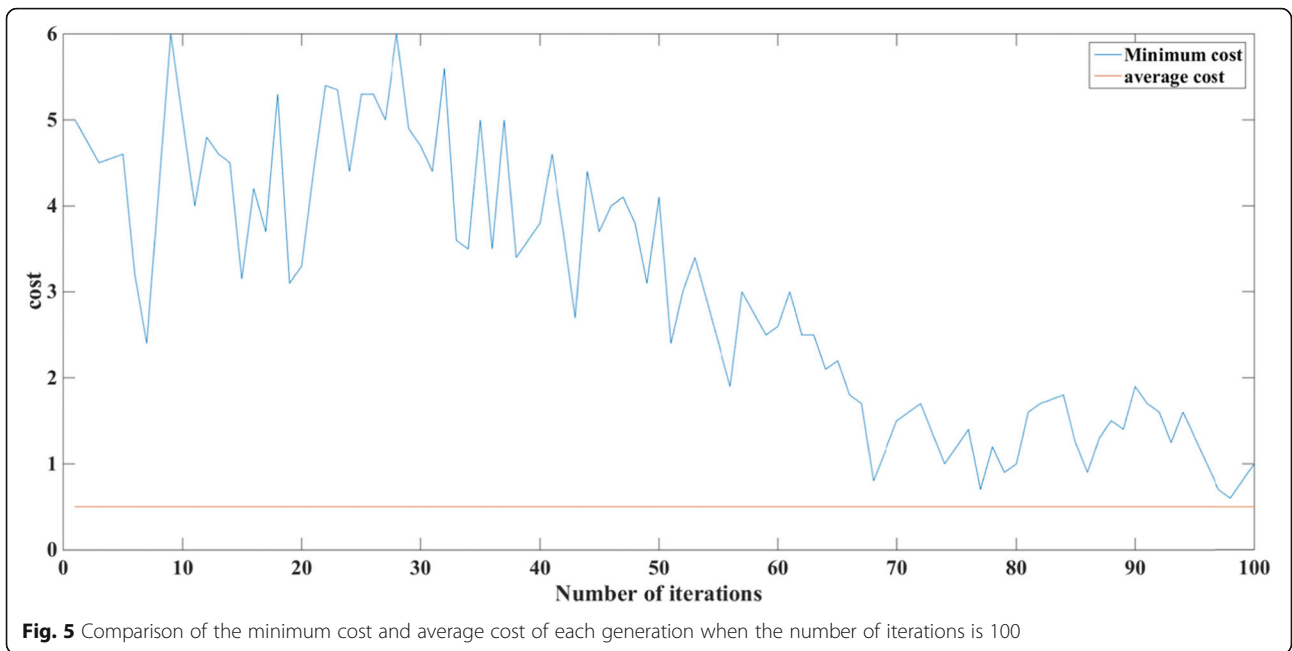
Set the maximum number of iterations in the MATLAB program to 10, 50, and 100. When the number of iterations is 10 or 50, the gap between the minimum cost and the average cost of each generation is relatively large. Figure 5 shows the comparison between the minimum cost and average cost of each generation when the number of iterations is 100. We can see that when the number of iterations is close to 100, the difference between the minimum cost and the average cost of each generation does not exceed 0.1, and given the maximum iteration with the increasing number of times, the average cost of cold chain logistics distribution is constantly approaching the minimum cost. The comparison of the minimum cost and average cost of each generation when the number of iterations is 100 is shown in Fig. 5.

### *Analysis of genetic algorithm optimization results*

#### *Analysis of the optimal solution of the optimization process.*

The optimization process of the genetic algorithm and the optimal solution in the iteration process are shown in Fig. 6. The line segment on the graph is a line

**Fig. 4** Comparison of optimization results

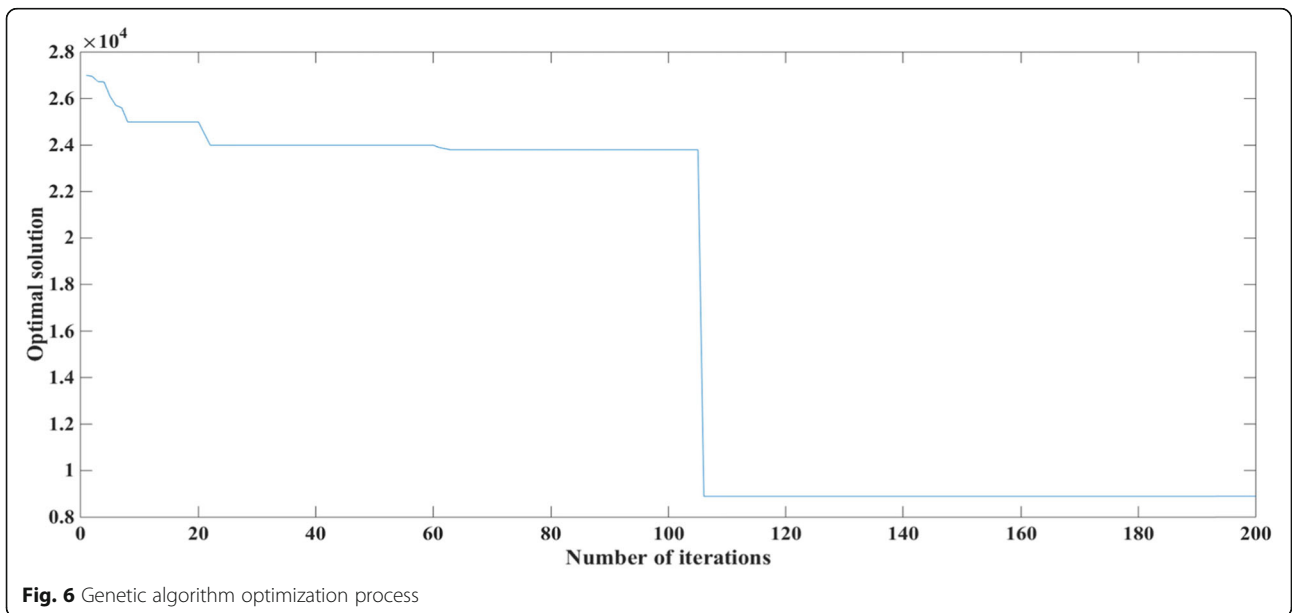


**Fig. 5** Comparison of the minimum cost and average cost of each generation when the number of iterations is 100

segment composed of 200 points, and each point is the result of each iteration. When the number of iterations reaches 105, the iteration results tend to be stable. In the actual running process, the number of termination iterations was tried at 200, 300, 500, 800, and 1000, and the final results were between 8000 and 9000. Therefore, this paper chooses 8150, with the lowest cost, as the approximate optimal solution. The genetic algorithm optimization process is shown in Fig. 6.

**Comparison of coarse-grained parallel genetic algorithm and classic genetic algorithm.**

Using 1, 2, 4, and 8 different numbers of processors, the execution times of the algorithm are 19.89, 14.52, 8.12, and 6.41, respectively. It can be seen that to a certain extent, the greater the number of processors is, the shorter the calculation time. The comparison results of different numbers of processors are shown in Table 2.



**Fig. 6** Genetic algorithm optimization process

**Table 2** Comparison of different numbers of processors

Number of processors	Execution time	Speedup (times)	Effectiveness
1	19.89	–	–
2	14.52	1.56	0.83
4	8.12	2.5	0.7
8	6.41	3.67	0.55

Depending on the degree of attention paid to vehicle delivery costs and delivery times, the time and cost of the calculated results are also different. Comparing the distribution time under the parallel genetic algorithm and the classic genetic algorithm, the experimental results are shown in Fig. 7. We can see that when the number of iterations is less than 50, the time difference between the two algorithms is not large, with both between 5.3 and 5.8. When the number of iterations is greater than 50, the fastest delivery time under the parallel genetic algorithm reaches 1.6 h, while the fastest delivery time of the classic genetic algorithm takes 3.25 h. The evolutionary algebra of the coarse-grained parallel genetic algorithm model and the evolutionary algebra of the classic genetic algorithm in the big data cloud computing environment are shown in Fig. 7.

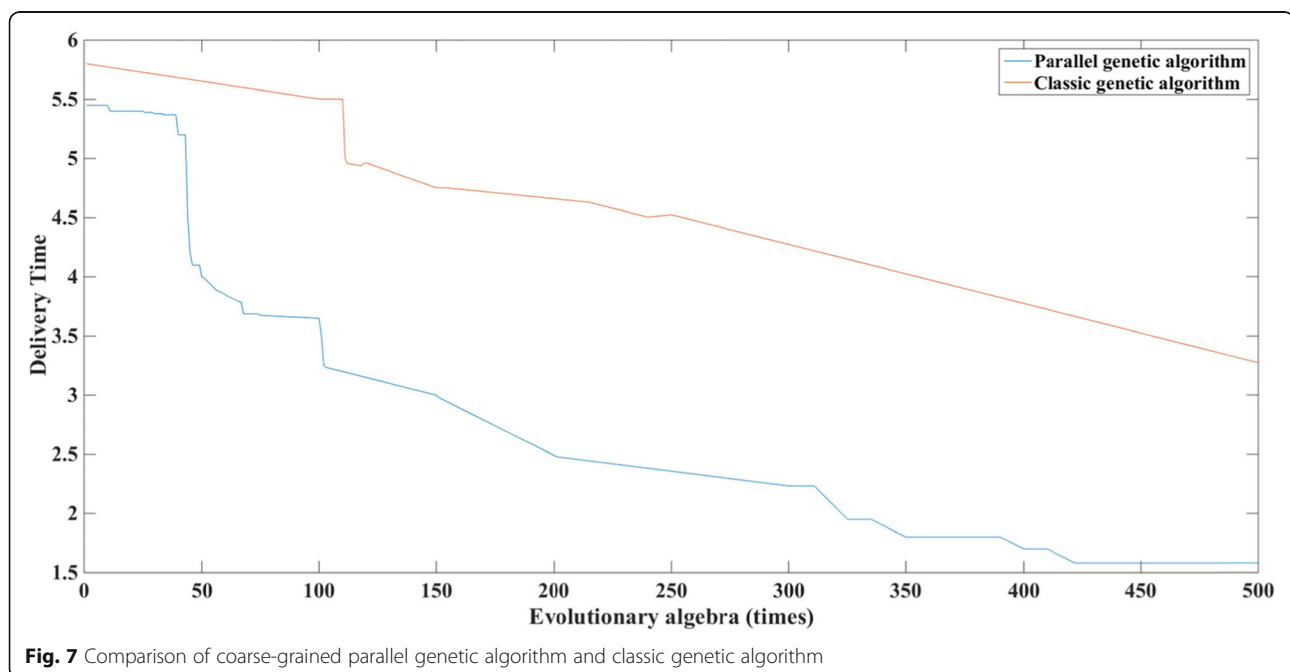
## Conclusions

(1) In the context of big data cloud computing, this paper studies the problem of the real-time optimization of cold chain logistics distribution vehicles. With cold chain logistics distribution vehicle optimization as the application background and big data cloud computing

as the technical support, real-time vehicle routing is optimized. Through the service model of cloud computing, cold chain logistics enterprises can complete the same calculation and processing process only by paying a low service fee, which reduces the cost and improves processing efficiency. Therefore, it is practical and feasible to study the cold chain logistics distribution vehicle routing problem in big data cloud computing.

(2) This paper analyses the parallel programming mode of big data cloud computing. To speed up the solution rate of real-time optimization models, the idea of parallel design of coarse-grained genetic algorithms in the big data cloud computing environment is given, analysis and simulation experiments of the algorithm are performed, and the experimental results verify the validity of the model.

(3) This article only analyses and explores the optimization of cold chain logistics distribution vehicle routing based on big data cloud computing. There are still many issues that need further research and improvement. 1) The application of big data cloud computing technology in the cold chain logistics industry has not been fully launched, it may be difficult to integrate information resources, and it requires considerable manpower and material resources to cooperate. 2) The multisource factors of cold chain logistics distribution vehicle routing optimization are analysed, the general method of analysis is given, and simulation experiments are performed, which need to be verified in a real big data cloud computing environment.

**Fig. 7** Comparison of coarse-grained parallel genetic algorithm and classic genetic algorithm

**Abbreviations**

IaaS: Infrastructure as a Service; PaaS: Platform as a Service; SaaS: Software as a Service

**Acknowledgements**

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

**About the authors**

Yi-hua Chen (1965.08--), male, the Han nationality, born in Shaoyang, Hunan, Master, Associate Professor, engaged in information technology research of the electronic & computer.

**Author's contributions**

All the authors take part in the discussion of the work described in this paper, did the experiments of the paper and Lin Shi wrote the paper. The author(s) read and approved the final manuscript.

**Funding**

This paper is the science and technology innovation project of Foshan city, Guangdong, R.China. Project Name: The electronic supervision platform of cold-chain transport based on mobile Internet technology. Project Number: 2016AG100472.

**Availability of data and materials**

We can provide the data.

**Ethics approval and consent to participate**

Approved.

**Consent for publication**

Approved.

**Competing interests**

These no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Received: 30 December 2019 Accepted: 29 April 2020

Published online: 13 July 2020

**References**

1. Lv Y, Duan Y, Kang W (2015) Traffic flow prediction with big data: a deep learning approach. *IEEE Trans Intell Transp Syst* 16(2):865–873
2. Boru D, Klizovich D, Granelli F (2015) Energy-efficient data replication in cloud computing datacenters. *Clust Comput* 18(1):385–402
3. Lailossa GW (2015) The new paradigm of cold chain management systems and it's logistics on tuna fishery sector in Indonesia. *Aal Bioflux* 8(3):381–389
4. Andreu-Perez J, Poon CCY, Merrifield RD (2015) Big data for health[J]. *IEEE J Biomed Health Inf* 19(4):1193–1208
5. Zaharia M, Xin RS, Wendell P (2016) Apache spark: a unified engine for big data processing. *Commun ACM* 59(11):56–65
6. Hashem IAT, Chang V, Anuar NB (2016) The role of big data in smart city. *Int J Inf Manag* 36(5):748–758
7. Liu K, Sheng W, Zhang D (2015) Big data application requirements and scenario analysis in smart distribution network. *Proc Csee* 35(2):287–293
8. Jiang D, Wu S, Chen G (2015) epic: an extensible and scalable system for processing Big Data. *Proc Vldb Endowment* 7(7):541–552
9. Zheng K, Yang Z, Zhang K (2016) Big data-driven optimization for mobile networks toward 5G. *IEEE Netw* 30(1):44–51
10. Zhou S, Xu Q, Qi Q (2016) Big data in mobile social networks: a QoE-oriented framework. *IEEE Netw* 30(1):52–57
11. Bera S, Misra S, Rodrigues JJPC (2015) Cloud computing applications for smart grid: a survey. *Parallel Distributed Syst IEEE Trans* 26(5):1477–1494
12. Mcgilvary GA, Barker A, Atkinson M (2015) Ad hoc cloud computing: from concept to realization. *Comput Sci* 265(11):1063–1068
13. Zhong Z, Chen K, Zhai X (2016) Virtual machine-based task scheduling algorithm in a cloud computing environment[J]. *Tsinghua Sci Technol* 21(6):660–667
14. Kaur T, Chana I (2015) Energy efficiency techniques in cloud computing- a survey and taxonomy[J]. *ACM Comput Surv* 48(2):1–46
15. Barsoum AF, M. (2015) Anwar Hasan. Provable multicopy dynamic data possession in cloud computing systems. *IEEE Trans Inf Forensics Security* 10(3):485–497
16. Cao Z, Lin J, Wan C (2017) Optimal cloud computing resource allocation for demand side Management in Smart Grid. *IEEE Trans Smart Grid* 8(4):1943–1955
17. Shen J, Member IEEE (2018) Anonymous and traceable group data sharing in cloud computing. *IEEE Trans Inf Forensics Security* 13(4):912–925
18. Sabi HM, Uzoka F-ME, Langmia K (2016) Conceptualizing a model for adoption of cloud computing in education. *Int J Inf Manag* 36(2):183–191
19. Wang J, Chen X, Huang X (2015) Verifiable auditing for outsourced database in cloud computing. *IEEE Trans Comput* 64(11):1
20. Cai H, Xu B, Jiang L (2017) IoT-based big data storage systems in cloud computing: perspectives and challenges. *IEEE Internet Things J* 4(1):75–87
21. Bo CHEN, Ao ZHOU (2016) An analysis of the cold chain logistics concerning agricultural products in Chongqing City. *Asian Agric Res* 08(1):19–21
22. Zhang H, Qiu B, Zhang K (2017) A new risk assessment model for agricultural products cold chain logistics. *Ind Manag Data Syst* 117(9):1800–1816
23. Ebing L, Mintian Z (2015) The new Gsp pharmaceutical cold chain logistics present situation and the suggestions of the development of quality management. *J Liaoning Univ Tradit Chin Med* 18(4):519–533
24. Xi-Chuan ZHAI, Yong-Feng LIU, Ye WANG (2015) Application of modeling techniques based on CFD in food cold chain logistics. *J Food Saf Qual* 1:207–211
25. Wang Y, Li J, Wang HH (2019) Cluster and cloud computing framework for scientific metrology in flow control. *Clust Comput* 22(1):1–10

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)