Risk Analysis of Big Data Based on Cloud Computing for the Inspection and Testing of Toxic and Hazardous Substances in Meat Products

Yajie WANG*, **Jinlin HE**, **Peng WANG**, **Jiao DAI**, **Bing YANG**, **Hong TAN**, **Guangcan TAO**
Guizhou Academy of Testing and Analysis, Guiyang 550001, China

Abstract Large-scale data emerge in food safety inspection and testing industry with the development of Internet technology in China. This paper was aimed at designing toxic and hazardous substance big data risk analysis algorithm in food safety inspection and testing based on cloud computing[1]. Cloud computing platform was set up to store the massive extensive data with geographical distribution, dynamic and high complexity from the Internet, and MapReduce[2] computational framework in cloud computing was applied to process and compute parallel data. The risk analysis results were obtained by analyzing 1000000 meat products testing data collected from the laboratory management information system based on web. The results show that food safety index $\overline{IFS} < 1$, which proves that the food safety state is in good condition.
Key words Cloud computing, MapReduce, Toxic and hazardous substances, Food safety risk analysis

1 Introduction

In recent years, with the rapid development of the Internet, cloud computing technology has been applied properly in the financial[3-5], automotive[6-7], power industries[8-9], *etc*. But there is no research presenting the combination of cloud computing and food safety risk analysis industry. Food safety state evaluation index system, proposed by Cong Li *et al.*[10], has been applied in food safety risk analysis. However, as the data for risk analysis are too small, the results are not convincing. Besides, the evaluation algorithms have not been applied under the cloud technology. Based on the Internet + food safety inspection and testing model driven in food industry, food safety inspection and testing data increase exponentially. At the same time, big data have been forming. Massive food inspection and data will undoubtedly bring a vast amount of information for the public, but the difficulty of detecting useful knowledge in massive food safety inspection and testing data in administration departments, enterprises and food testing institutions increases. Food safety risk analysis has always been a difficult problem in the food safety field, and the risk assessment algorithm based on cloud computing for food safety, toxic and hazardous substances can effectively solve the mining and utilization of potential value of massive data. This paper collects the food inspection and testing big data of national testing agencies from the laboratory information management system (LIMS) developed on web together, and makes full use of high availability and high degree of virtualization features on cloud computing platform[11] to dynamically schedule and allocate resources, thereby meeting the precise risk analysis and efficient data mining demand. This paper puts forward a risk analysis algorithm based on cloud computing for the inspection and testing of toxic and hazardous substances in food safety, the mathematical model is used in food safety inspection and testing. This model is a kind of algorithm application technology which lets food safety inspection and testing results reflect multivariate relationship between data, the algorithm has been applied and operated in food safety cloud platform in Guizhou Province, China, which has achieved good results. This paper is the practical application of Internet + food safety inspection and testing mode, and it realizes a risk communication method with massive data information as the center.

Received: June 16, 2017 Accepted: August 20, 2017
* Corresponding author. E-mail: wyj@gzata.cn

2 Materials and methods

2.1 Cloud computing Cloud computing[12] is a computing model for realizing the distributed large-scale data computing, through the gathering of the mass distribution food detection results from food safety inspection and testing laboratory resources in different places and in different time, providing storage and computing ability for massive food data. Cloud computing makes use of platform virtualization and dynamic resource allocation techniques to provide data on-demand services for the administrative authorities, the food industry enterprises, testing agencies, the media and the public. It avoids the waste of resources and improves the utilization frequency and application performance of cloud server. In general, cloud computing uses a large number of inexpensive and interconnected computers on the Internet for task processing in demand, as a result of providing big data storage resources and computing resources for food safety inspection and testing results. Users do not need to understand the management of cloud computing infrastructure. At the same time, cloud computing provides the ability of dynamic load balance and horizontal expansion. With the increase of the amount of data, more data nodes for computing are required. Cloud computing adds nodes to the data center at runtime and automatically transfers some of the load to the new node to calculate. And the mechanism of cloud computing maintains the load balance between nodes, and it improves the load capacity of the business[13]. In addition, making use of virtualization technology, cloud computing technology can achieve distributed computing, centralized management of information resources and intelligent deployment. There are three main types of mainstream business model of cloud computing, namely Platform as a Service

(PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS). The logic diagram according to this paper is as shown in Fig. 1. In cloud computing environment, all data resources are divided into multiple data centers. A data center has tens of thousands of data nodes, these data nodes are efficiently connected through the Internet, which can provide users with storage and computing resources. With the rapid development of cloud computing, virtualization can be applied to achieve low-cost and large-scale computing platform. Virtualization technology can be applied to the effective use of idle computing platform. It abstracts the idle computing resources which contribute to the formation of independent virtual server instances, which, therefore, independently assigns data processing and calculation and realizes the virtualization of the underlying hardware.

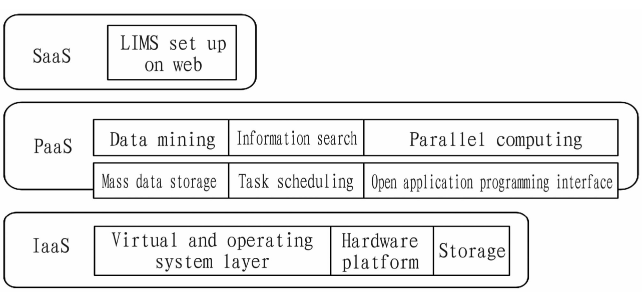

Fig. 1 Cloud computing logic diagram

2.2 Parallel computing model in cloud computing environment MapReduce[14] is a distributed parallel computing model framework for processing large-scale data in a cloud computing environment, which is proposed by Google laboratory. MapReduce processing is divided into two steps, namely Map step and Reduce step. Tasks are formed by a large number of independent and parallel Map tasks and Reduce tasks. Each step can be indicated by a function, namely Map and Reduce functions. Besides, each function uses a key <key, value> as input and output. Map steps divide large-scale data processing operations into various small independent tasks, and at the same time, the Map function will enter the key <key, value>, utilizing virtualization technology to allocate multiple sub-tasks to the computing nodes in the computer clusters and do distributed processing. The Reduce function processes the value with the same key, and then generates key pairs. The Reduce step summarizes processing results of various tasks, finally it will output the analysis results. In the development process of MapReduce parallel computing model, firstly, through reading the massive data collected from the Internet applications, the required processing and calculating data are split into a lot of data blocks. In the following steps, users write the corresponding MapReduce implementation procedures. Through the operation of distributed computer cluster, users can decompose the large data set, thereby forming mass key < key, value > pairs based on the formation of series of data fragments. Such a process can be implemented by Map and Reduce interface. The results generated by the MapReduce programming model are finally saved in HDFS. According to MapReduce framework, cloud computing can achieve large-scale data processing and analysis. When using the MapReduce computing framework for parallel computing, the large-scale data should be consistent with the structure. For massive data mining, it is usually associated with the complexity of the calculation algorithm such as the approximate solution, data dimension reduction and program iteration. And as the industry is different, data mining algorithm often needs to be customized according to the industry data. Therefore, the massive data mining technology based on cloud computing and specific industry has attracted the attention of the academic community and the industry. The amount of data is usually able to reach GB or TB level in the present. Cloud computing platform uses HDFS to support large scale data, which can be extended in the large-scale distributed computing cluster nodes and provides wide data bandwidth.

2.3 Risk analysis of big data based on cloud computing for the inspection and testing of toxic and hazardous substances in meat products The current big data analysis techniques based on cloud computing have been applied to the Internet industry and software architecture development, software design, function realization, which can fully expresses the characteristics of the Internet big data and business logic. When it is applied to the specific industry in addition to the Internet, we first need to analyze the characteristics of the large data of specific industries, and then customize development based on business logic. In order to meet the deep analysis and mining requirements of food inspection and testing of big data according to the industry requirements and data characteristics, this study firstly collects the Internet + food safety inspection and testing laboratory big data based on cloud platform through food safety LIMS deployment on website. Secondly, based on the Extract-Transform-Load (ETL) of data, combined with the latest developments in the industry and the actual deployment experience of cloud computing technology, this paper presents the designing of big data risk analysis according to toxic and harmful substances in food safety in the PaaS layer (Fig. 1). In the following step, the study implements the MapReduce computing framework based on cloud computing for distributed computing and analysis to obtain the final analysis result which can be visualized in software system based on the Internet in the SaaS layer (Fig. 1).

2.3.1 Data storage. In order to ensure the highly reliable application of food safety inspection and testing data, this paper adopts the distributed storage mode of cloud computing to store the data. According to food safety inspection and testing data collection of different laboratories in different time and different locations in LIMS on the Internet, the designed mechanism storage data are in multiple copies. The specific operation is dividing the inspection and testing results into different data block size, rather than the basic copy. Then the mechanism stores data blocks in the corresponding data nodes based on different configuration information.

2.3.2 Computing to store migration. In this paper, the research related to risk analysis of the detection of food safety, toxic and hazardous substance big data based on cloud computing is the research of a data mining algorithm in food industry. The data mining algorithm designed in this paper is based on the data of massive testing results. In the distributed system, the network bandwidth resources are relatively scarce, and the computing power of each node is relatively rich. In the specific operation process of task scheduling, the cloud computing platform integrates the storage nodes and computing nodes together, and allocates and executes specific tasks in the distributed storage nodes which store corresponding input data blocks. This mechanism allows parallel tasks to read input data on the local computer node, which can reduce network data flow to a certain extent[15].

2.3.3 Data mining. Data mining, also known as knowledge discovery process, is the process of discovering potentially understandable and valuable information from the large scale data through the mining algorithm. The risk analysis of food safety toxic and hazardous substances is a concrete realization of data mining, the ultimate goal of which is to extract the knowledge from the mass inspection and testing data to extract the potential value. We hope that the data scale will become bigger, so that the results of the data mining will become more accurate and the risk analysis will be more convincing. Such risk analysis requires high levels of environment for the development and application. In cloud computing platform, massive data are stored in the data center, cloud computing makes use of specific application risk analysis algorithm to dynamically allocate resource to compute. And it uses the fault tolerance mechanism to ensure the reliability and extensible property of data mining algorithms. Based on the above ideas, this paper designs the data mining system based on cloud computing technology. The risk analysis algorithm for toxic and harmful substances in food safety is a kind of data mining algorithm library. The whole structure is shown in Fig. 2. In the data mining system designed in this paper, the nodes are divided into the main control node (TotalCtrlNode) and the working node (WorkerNode). There is only one TotalCtrlNode in the whole system structure, composed of NameNode, SecondaryNameNode, data warehouse in HDFS, Jobtracker in MapReduce calculation model, as well as the data mining algorithm. The risk analysis algorithm designed in this paper is a kind of data mining algorithm in the library. Along with the business system, more risk analysis algorithms based on food safety industry can be designed in the data mining algorithm library in the future. Among them, NameNode is the main server of HDFS which is used for metadata of file system management and storage. It also implements the namespace operation of HDFS such as open, close and rename, and divides the file into several blocks, then maps into the DataNode in WorkerNode. The data mining system designed in this study can deploy a number of WorkerNode nodes in demand, which is composed of DataNode in HDFS and TaskTracker in MapReduce calculation model. The DataNode and TaskTracker are responsible for data storage and computing respectively. The actual large-scale data are separated into data blocks stored in DataNode. DataNode is also responsible for the actual implementing of read and write requests in accordance with the NameNode command, and the performing of data blocks' creating, copy and deleting, etc. Meanwhile, TaskTracker utilizes the risk analysis algorithm developed in this paper to deal with the actual data mining of various data blocks. In the actual operation of the system, we firstly acquired data with different locations, different time and different laboratory from the application system LIMS based on web, and saved them into data warehouse. The NameNode in HDFS based on cloud computing model will automatically divide the data files into several pieces and store them in each DataNode in the WorkerNode. SeondaryNameNode and NameNode are deployed and run on different machines. The main function of SeondaryNameNode is to assist with image files and transaction log processing by NameNode. The developed risk analysis algorithm is put into the algorithm library of TotalCtrlNode. The JobTracker under MapReduce computing model in TotalCtrlNode node will schedule and coordinate tasks between computing nodes, and distribute and monitor the data mining tasks across the WorkerNode node according to the requirements. And it will reexecute the failure task. TaskTracker will call these tasks on-demand, or combine the local database to perform tasks assigned by TotalCtrlNode according to the specific business requirements. Besides, TaskTracker will give the feedback of the computing results and corresponding state information to TotalCtrlNode. JobTracker and TaskTracker adopt Master/Slave model to work, Master and Slave send mutual commands to achieve interaction. The structure of this mechanism has high fault tolerance. In addition, it reperforms the computational tasks of broken WorkerNode. At the same time, migration information is sent to TotalCtrlNode, which will redistribute tasks to the other WorkerNode nodes. The system storage is comprised by NameNode, SecondaryNameNode, DataNode in HDFS and data warehouse of LIMS on the Internet terminal while the computing components is constituted by JobTracker, TaskTracker under MapReduce model and the designed data mining algorithm library in this paper. The data mining system based on cloud computing technology uses the Hadoop as the big data solution, making use of the computer clusters deployed in the cloud and based on the Linux operating system as the hardware devices. In this paper, the data mining algorithm is designed in the data mining system based on cloud computing to detect the risk from food inspection and testing of toxic and hazardous substance big data. Because of the various factors affecting food safety and extensive content, the risk analysis of the testing results needs to start from a global perspective, considering the overall status of food safety from different levels. The principle of risk analysis is utilizing the connection of the important degree of food safety state, the change of food safety state in coordination with risk analysis to evaluate. It also uses reliability and sensitivity which reflects the state change of food safety to assess[17].

(i) Technology principles of threshold method application. The theoretical basis of the structure calculation of food safety toxic and harmful substances inspection and testing big data is that there is reasonable and comparable threshold value between human body and the animals[18]. However, people's sensitivity is higher than that of animals. In addition, there is large genetic difference, and the dietary habits are more different. In view of this, ADI[19], PTWI[20] or RDI[21] can be used to judge. (ii) Risk analysis of big data sets for toxic and hazardous substances in food safety inspection and testing. Risk analysis involves specific testing index data. Taking the BHC($C_6H_6Cl_6$) content in vegetables and the formaldehyde content in beer for example, the testing results of big data are the basic data of the risk analysis. To determine the risk analysis of large data sets, it is according to the domestic and foreign standards or regulations of all relevant food and based on the toxic and harmful

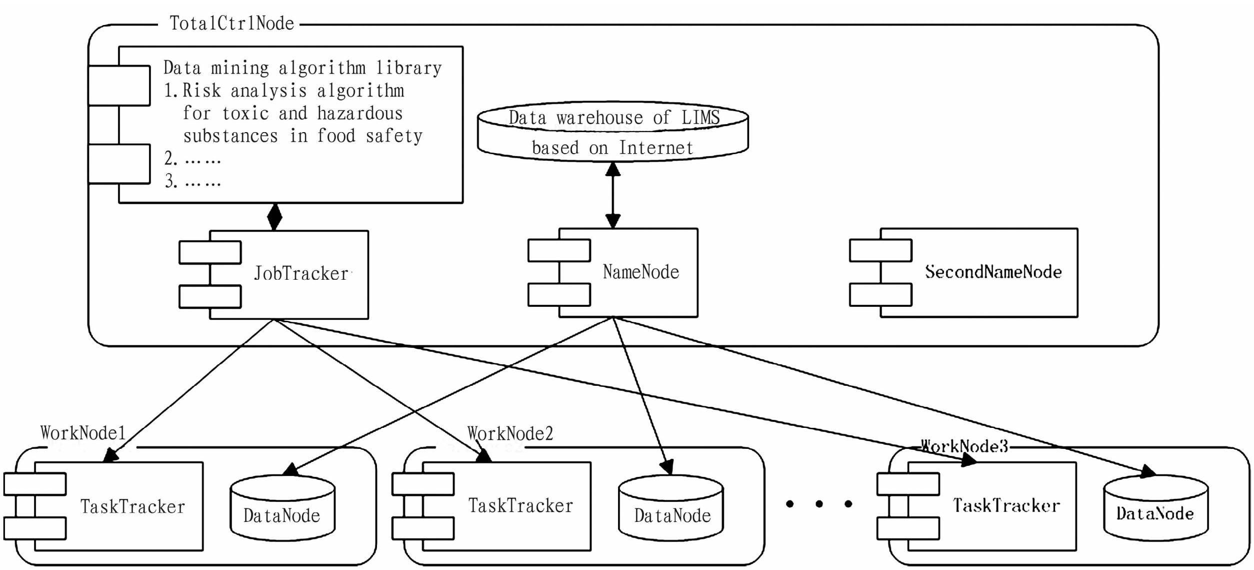

Fig. 2 Architecture of data mining system based on cloud computing

indicators and related testing big data across regions and laboratories collected by LIMS on web. This paper designs and implements the following algorithms based on the use of the collected original data from the Internet and cloud computing technology to compute relevant parameters, thereby doing further risk analysis. (iii) Dimensionality reduction of multidimensional factors affecting food safety. According to the completeness and minimum principle of food safety inspection and testing, first of all, we reduce the dimensionality of factors which affect the food safety of reduction into the following several factors[22]. Then we implement the relevant calculation, these five factors are namely the risk evaluation index. (a) The average content (AVE). In this paper, the average content of a certain type of toxic and hazardous substance in a specific food is namely in the average content(AVE_{item}), also the average value of a specific test of toxic and hazardous substances. The equation is as follows:

$$AVE_{item} = \frac{\sum_{i=1}^{n} C_i}{n} \quad (1)$$

where AVE_{item} is the average value of the testing index data, computed by the toxic and harmful substances' testing results; C_i stands for the content of testing result in sample i; n is total number of samples. (b) Limited standards (STA). Generally speaking, for each testing result, in accordance with international or national standards, the standard will be specified in the limited standards (AVE_{item}). In the field of food safety, risk analysis of toxic and harmful substances, pathogenic microorganisms and other harmful substances is usually concerned, therefore limited standards (AVE_{item}) for the risk analysis of food safety inspection and testing should be the upper limit. The standard regulated for "no" or "not tested" and other non-digital situation, it can be regarded as the "0". (c) Overrun rate(OUT). Overrun rate(OUT_{item}) is in the entire data set, the number of the testing samples' results which exceed the limited standard (STA_{item}). The algorithms can be seen as equation (2):

$$OUT_{item} = \frac{M}{N} \times 100\% \quad (2)$$

where OUT_{item} stands for the overrun rate; M is the number of data beyond the limited standard; N is the number of the entire data set. (d) Overrun degree (OUD). Overrun degree represents the whole monitoring structure, all overrun data beyond the limit of the standard (STA_{item}).

$$OUD_{item} = \frac{\sqrt{\sum_{j=1,2,\cdots,m} (C_j - STA_{item})^2}}{M} \quad (3)$$

where OUD_{item} represents the overrun degree; C_j is a specific data value which exceeds the limit; STA_{item} is the index of a limited standard; M represents the number of data beyond the standard limit. (e) Maximum (MAX). In the whole data set, the maximum value of the inspection results is in the whole data set. In this paper, we designed five evaluation indexes for specific test items in a specific food, and solved the problem of the risk analysis of the specific items in the specific food. Because the food composition is very complex, which may contain toxic and harmful substances and a variety of pathogenic microorganisms, to realize the risk analysis for food inspection and testing, we need to establish the basic food category index. The risk analysis of the index of food category is mainly in the following two aspects. (a) The qualified rate of food (σ). For a specific food, such as meat product, according to the specific structured data results of each testing item, we can evaluate whether the meat samples reach the national standard by comparing the results with national standard, namely whether it is qualified. Different kinds of food conformity assessment standards and methods are different. For example, in some kinds of food, as long as there is any structured testing result of the item beyond its limits, it can be judged as the unqualified food. In some foods, you need to consider a number of comprehensive testing items, and then evaluate. Either way, all the samples can be evaluated as qualified or not. Thus, we can get the qualified rate of the food (σ_{food}) in the whole cloud data set as follows:

$$\sigma_{\text{food}} = \frac{L}{N} \times 100\% \quad (4)$$

where σ_{food} stands for the food qualified rate; represents the sample volume in the entire data set to meet national standards, namely the amount of qualified samples; N illustrates the total sample volume in data set. Thus, $(1-\sigma_{\text{food}})$ represents the unqualified rate of food. (b) Food insecurity degree (h). Food insecurity degree (h) is used to represent the overall level of the insecurity of a food, and it is a dimensionless number. Because there are many food substances such as microbial or chemical substances which harm food safety, a structure to represent the degree of unsafe food is required. We need to do the dimensionless process to the hazardous substances in food, at the same time, also need to know the harm coefficient[22] of harmful substances, the formula of food insecurity degree (h_{food}) is as follows:

$$h_{food} = \sum_{AVE_{item} > STA_{item}} \left(\frac{AVE_{item} - STA_{item}}{STA_{item}} \times F_{item} \right) \quad (5)$$

where h_{food} is the unsafe degree of a certain kind of food; F_{item} stands for the risk factors of the toxic or hazardous substances or pathogenic microorganisms. Typically, $F_{item} \leqslant 1$, AVE_{item} is the average of the testing index value when STA_{item} stands for the limited standard of the index. The equation (5) illustrates that in all the hazards in the food, we do the dimensionless process to all the hazardous substances beyond the limit, $\frac{AVE_{item} - STA_{item}}{STA_{item}}$ namely. Then, we use it to multiply F_{item}. Finally, we can get h_{food} by cumulative summation. (iv) Risk analysis of overall state data in food safety inspection and testing. The risk analysis of overall state data in food safety inspection and testing is based on the food safety index according to the method of exposure assessment (IFS). When there is hazardous substance in food, IFS can determine whether the consumers are healthy, whether there are hazards. That is, IFS can assess the overall state of food safety.

$$IFS = \frac{\sum_{C=C_1,\ldots,C_n}^{n} IFS_C}{n} \quad (6)$$

where IFS_c represents the food safety index of some hazardous substances; C in food influences the health of consumers; n illustrates the total amount of samples. Because the toxic effects of food hazards are related to the absolute amount the food hazards entering into the human's body, using the principle of threshold approach application, when designing the evaluation of poisonous and harmful substance of material in food safety, we believe that to evaluate the overall state in food safety, it is more scientific and reasonable to make use of the actual intake of hazardous substances by the human body through the diet divided by the safe intake. The safety factor for long-term animal testing data is 100[18], but different countries adopt different standards, generally using ADI[19], PTW[20], or RDI[21] data. Under the guidance of such a theory, first of all, it can be used to evaluate the food safety index IFS_C, which is a kind of food safety index that influences health of the consumers, affected by hazardous substance C in certain food. The calculation formula is as follows:

$$IFS_C = \frac{EDI_C \cdot f}{SI_C \cdot bw} \quad (7)$$

where SI_C stands for the intake of food safety, it can be valued by the ADI, PTWI and RDI standard of different hazardous substance; bw represents the average weight (kg), the default value is 60; f illustrates the correction factor, if the safe intake uses ADI, PTDI, the value of f is 1, and if the safe intake uses PTWI and other weekly intake data, the value of f is 7; EDI_C represents the estimated value of actual daily intake of chemical substance C, EDI_C can be calculated by equation (8):

$$EDI_C = \sum (R_i \cdot F_i) \quad (8)$$

where R_i stands for the residual level C in food i, namely the testing result of chemical substance C, the unit is mg/kg; F_i shows the estimated daily consumption of food i, the unit of which is g/(person · day). Equation (7) can be used to calculate the food safety index IFS_C of hazardous substance C, we can conclude the impact of the hazardous substances on food safety by the above equation. When $IFS_C < 1$, the hazard subtance C has no impact on food safety; when $IFS_C \leqslant 1$, the impacts of risk of hazard subtance C on food safety is acceptable; when $IFS_C > 1$, the impacts of risk of hazard subtance C on food safety is beyond the acceptable limit. Taking into account the eating habits of the consumer and a variety of IFS_C food and chemical residues in the case, the value ofowns the additive property:

$$\overline{IFS} = \frac{\sum_{C=C_1,\cdots,C_n}^{n} IFS_C}{n} \quad (9)$$

In this situation, when $\overline{IFS} < 1$, it is concluded that food safety is in good condition and it is acceptable; when $\overline{IFS} > 1$, the food safety state of the consumers is not acceptable. No matter what kind of the above situation, for any $IFS_c > 1$, it is explained that the exposure of chemical substance C has exceeded the acceptable level, and we should enter the risk management process.

3 Results and discussions

Data mining based on cloud computing technology is basically consistent with the process of traditional data mining. The system is made up of 6 physical computers, in which, 1 computer is the TotalCtrlNode, the other 5 are WorkerNode. Based on the LIMS application system on web, we collect 1000000 testing data of the meat product of local brand in Guizhou Province in different time, different location and different laboratories in September. And we do the risk analysis based on the testing result of food contaminants Pb, Hg, and As. We implement data mining and analysis, in order to look forward to doing the big data risk analysis and assessment of toxic and hazardous substances in food safety inspection and testing. The acquired data can be firstly saved in the data warehouse of TotalCtrlNode. Then NameNode in HDFS will divide the data into different blocks, and save the data blocks into the DataNode in 5 WorkerNode. Based on the food safety inspection and testing toxic substance, the risk analysis algorithm is designed in the data mining algorithm layer in the TotalCtrlNode node, using JobTracker in MapReduce and Cloud computing to analyze and compute different data blocks, and using TaskTracker to implement the analysis and computing tasks. The acquired results will be stored into various DataNode. According to the national standard GB2762 in China, the limited standard STA of contaminants Pb, Hg, As is 0.5mg/kg, 0.05mg/kg and 0.5mg/kg[23]. In 1000000 groups of testing data, there are three unqualified groups in the testing result of Pb and Hg, two unqualified groups in the testing result of As. The unqualified indicator testing results are as follows: Pb (0.6 mg/kg, 0.65 mg/kg, 0.66 mg/kg); Hg (0.06 mg/kg, 0.068 mg/kg); As (0.6 mg/kg, 0.611 mg/kg). In the application example, this paper utilizes the intake data of ADI[19], theof Pb, Hg and As is 0.00356 mg/kg, 0.0001 mg/kg and 0.00215 mg/kg respectively. The average daily consumption of this meat product is 100g, namely, 0.1kg. Based on the big data risk analysis algorithm of food safety toxic and hazardous substances designed in this paper, we implement the cloud computing according to the collected testing data of the meat product, and the risk analysis results in Table 1 are obtained. When calculating the unsafety degree H_{food}, because the acquired $AVE_{item} < STA_{item}$, there is no unsafety degree. As the $\overline{IFS}$ of the contaminants Pb, Hg and As < 1, we can draw the conclusion that the consumers' food safety state is good.

Table 1 Big data risk analysis results of the safety of poisonous and harmful substance in meat products

Items	Pb	Hg	As
AVE_{item} //(mg/kg)	0.069	0.022	0.008
STA_{item} //(mg/kg)	0.500	0.050	0.500
OUT_{item}	0.003	0.002	0.002
OUD_{item}	0.080	0.010	0.007
MAX_{item} //(mg/kg)	0.660	0.068	0.611
σ_{food}	0.997	0.998	0.998
$1-\sigma_{food}$	0.003	0.002	0.002
h_{food}	–	–	–
$\overline{IFS}$	0.0324	0.3683	0.006

4 Conclusions

This paper designs a risk analysis algorithm for food safety toxic and harmful substances big data based on cloud computing in the PaaS layer of the cloud platform through the acquisition of Internet-based application system LIMS, the algorithm is designed to achieve MapReduce programming model based on cloud technology. Compared with the risk analysis of traditional food safety, the use of cloud computing model makes it cheaper, faster and easier to increase computing nodes, and due to the high tolerance of cloud computing, the data loss because of the node damage can be avoided. In this paper, efficient analysis and value mining of the food safety inspection and testing data collected by different location, different laboratories and different testing organizations are realized. It is a kind of risk communication method based on the mass data information and cloud computing. With the increase of the amount of data, the accuracy of the data mining results will be improved, the analysis results will be more valuable, and the advantage of risk analysis algorithm will become more obvious.

References

[1] MELL PM, GRANCE T. SP 800 – 145. The NIST definition of cloud computing [M]. National Institute of Standards & Technology, 2011.
[2] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107 – 113.
[3] XIE SQ. Discussion on cloud computing and its application in financial field [J]. Finance and Economy, 2010(11): 9 – 11. (in Chinese).
[4] YANG DQ. The case study of ERP implementation on company E [D]. Shanghai: Shanghai Jiao Tong University, 2013. (in Chinese).
[5] XIAO SR, XIAO KS. On the financial innovation of cloud computing era [J]. The Banker, 2013(3):116 – 119. (in Chinese).
[6] CAO ZJ, LIN J, SONG YH. Optimization model for resources allocation of cloud computations in active distribution networks [J]. Proceedings of the CSEE, 2014, 34(19):3043 – 3049. (in Chinese).
[7] LI SQ. On the data platform design of automotive aftermarket based on Hadoop [D]. Guangzhou: South China University of Technology, 2013. (in Chinese).
[8] ZHANG LX, XIN XJ, CUI B, *et al.* Research on Automotive Product Design Oriented Virtual Reality Service Platform [J]. Journal of System Simulation, 2014, 26(10):2407 – 2411. (in Chinese).
[9] WU KF, LIU WT, LI YH, *et al.* Cloud-Computing Based Power Big Data Analysis Technology and Its Application [J]. Electric Power, 2015, 48(2):111 – 116. (in Chinese).
[10] LI C. Food safety monitoring and warning system [M]. Beijing: Chemical Industry Press, 2006. (in Chinese).
[11] ANG Q, VARELA CA. Impact of cloud computing virtualization strategies on workloads' performance [C] IEEE International Conference on Utility & Cloud Computing. IEEE Computer Society, 2011:130 – 137.
[12] BUYYA R, YEO CS, VENUGOPAL S, *et al.* Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility [J]. Future Generation Computer Systems, 2009, 25(6):599 – 616.
[13] ARMBRUST MF, *et al.* Above the clouds: A berkeley view of cloud computing [J]. Eecs Department University of California Berkeley, 2015, 53(4):50 – 58.
[14] DEAN J, GHEMAWAT S. MapReduce: Simplified data processing on large clusters. [J]. Communications of the ACM, 2008, 51(1):107 – 113.
[15] YING Y, REN K, LIU ZT. Data mining based on cloud-computing technology [J]. Microelectronics & Computer, 2013, 30(2): 161 – 164. (in Chinese).
[16] HE Y, WANG WQ, XUE F. Study of massive data mining based on cloud computing [J]. Computer Technology and Development, 2013, 23(2):69 – 72. (in Chinese).
[17] HUANG XJ, LIU BL. Warning index system for food security [J]. Journal of Harbin University of Commerce: Natural Sciences Edition, 2008, 24(5):621 – 623. (in Chinese).
[18] JIANG DG. An empirical study on lead, cadmium pollution and dietary exposure assessment [D]. Chinese Center For Disease Control and Prevention, 2008. (in Chinese).
[19] LU FC. A Review of the acceptable daily intakes of pesticides assessed by WHO [J]. Regulatory Toxicology & Pharmacology, 1995, 21(3): 352 – 364.
[20] HERRMAN JL, YOUNES M. Background to the ADI/TDI/PTWI [J]. Regulatory Toxicology & Pharmacology, 1999, 30(2):109 – 113.
[21] HEGSTED DM. Recommended dietary intakes of elderly subjects [J]. American Journal of Clinical Nutrition, 1989, 50(5 Suppl): 1231 – 1235.
[22] LIU WJ. Study on food safety control in China [D]. Xianyang: North West Agriculture and Forestry University, 2006. (in Chinese).
[23] TIAN HY, REN XM, MA W, *et al.* The interpretation on national food safety standards in food contaminants (GB2762 – 2012) [J]. Beer Science and Technology, 2013 (25): 295 – 296. (in Chinese).