International Journal of Database Theory and Application
Vol.8, No.5 (2015), pp.293-304
http://dx.doi.org/10.14257/ijdta.2015.8.5.26

# Research on the Performance Optimization of Hadoop in Big Data Environment

Jia Min-Zheng

*Department of Information Engineering, Beijing Polytechnic College, Beijing China 100042*
*jmz@bgy.org.cn*

## *Abstract*

*In the age of Internet, the data transmission and storage got rapid progress, however, data processing and information extraction is still exist many problems to solve. Under the condition of so much data, processing data, get useful information; In cloud computing, big data environment to adopt the method of distributed computing, such a large complex networks, however, requires a simulation environment, for comparison and optimization platform, it can save development costs. Hadoop can evaluate the performance of distributed cloud computing platform, so the Hadoop performance directly affects the evaluation on the performance of the big data cloud computing, which fully show the importance of performance of Hadoop. Algorithm is improved based on Hadoop platform, using the particle swarm optimization algorithm improved the calculation and implementation of the Hadoop platform, so as to improve its ability to execute and compute, the calculation results and analysis show that the proposed scheme is effective.*

***Keywords:*** *Hadoop, Particle swarm algorithm; Hadoop execution algorithm; Calculate the clustering*

## 1. Introduction

With the rise of the Internet, it enables the data on a global scale to rapid transmission of data, especially the rapid development of optical fiber communication and cellular wireless access technology innovation, human anywhere in a more quick and convenient means of access to the Internet, around the various service is booming of Internet, mobile Internet services, in particular, it is varied and retrofit. At the same time, the Internet is booming, making access nodes not only including computer, also includes the data transmission by the sensors in the object and all of this will cause a problem, that is huge access nodes, through the Internet collect and input the mobile Internet, it has the generation of data stream, how to analyze such data and get useful information, so as to guide the human production and life in the future. The data selection calculation and screening promoted greatly development of parallel computing method. Due to network access nodes is too many, and with the development of the hardware, so the computing power of the individual nodes is improving. So we can solve the problem of data processing and filtering, along with the development of the data processing of parallel computing, the scholar proposed cloud computing method, it is different from the traditional parallel computing. It can accord the topological structure of network and computing tasks, assigned to computing tasks of each node, which can quickly get the calculation results. Say that cloud computing promoted the network era to the era of big data, through the development of the era of big data, humans can within the scope of the most widely, with the method of data processing to get the information

ISSN: 2005-4270 IJDTA

they want to know. In the process of cloud computing, we need to consider the speed of cloud computing and analyzing the data accuracy, in the Internet now, however, we use the new cloud computing method to validate the performance of cloud computing method is likely to run now network transmission of damage, and the influence of transmission performance. And using the interconnection network resources to call at the same time, it will cause a lot of unnecessary spending. In order to avoid unnecessary waste of manpower and material resources, so it is best to computer simulation approach to performance evaluation of cloud computing algorithm, it is convent to make comparison and evaluate the effectiveness of the algorithm.

However, what the cloud computing platform to simulate? In Google company proposed the operation of the platform that based on MapReduce at the beginning of this century, the platform through the coupling way of the business and the application, then it can be used to simulate the cloud computing model of distributed computing platform, and through constant development, eventually it became the most sought-after simulation evaluation calculation method of the cloud computing platform. With the advent of the era of big data, the human in the field of various disciplines can get an unprecedented amount of data; it is possible to find a more useful rule, which requires the existing big data filtering, statistics and analysis. This involves artificial intelligence data mining algorithms, this algorithm can be implemented through computer database correlation rule. In numerous association rules, however, there is connection and similarity, how to contact and similarity of the existence of classification and summary, it is the first step to realize finding a rule, generally referred to as data clustering process. However, clustering needs a variety of methods, such as some bionics methods, genetic algorithm, ant colony algorithm, *etc.*, all these methods have advantages, but its computational complexity is generally higher, how to improve the execution speed of the data mining method and reduce the computational complexity is always the core problem of big data analysis. Yet for Hadoop cloud computing platform, its business and understand the coupling calculation procedures, so you can separate the business and the calculating program to consider. Business and the application aspect, however, the application problem is the core of the problem, these problems always determines the authenticity of the computing speed of the entire platform and simulation. So aiming at this problem, this research mainly studies the application of data on the clustering algorithm, in data algorithm, particle swarm method with its recent calculation speed, low computing complexity and become the focus of research. Particle swarm method also has its own disadvantages, however, is due to the initialization value choice is different, it’s easy to get not calculate the global optimal value clustering results. Aiming at this problem, this study adopts genetic algorithm in the choice of initial value advantage, integrated the advantages of two methods, an improved particle swarm optimization (PSO) algorithm; this method effectively improves the computation efficiency and accuracy of the algorithm.

For the processing of data in Hadoop, however, have accumulated a certain amount of data sampling value, and through the analysis of the data sampling values, find the data of characteristic parameters of screening and analysis of the scope of data parameters, can get a characterization parameter of the characteristics of the data, according to the characteristics of the data of relational parameters and clustering, to find the relationship between each other. These relationships through the analysis of the artificial intelligence, can become the human need to gain knowledge, this undoubtedly accelerate the human to know the world and the nature of the energy and speed. So the study of this topic is original and great practical significance.

In Hadoop, calculation as the core does not shake, computing capability directly affects the evaluation of cloud computing, the process of execution is the evaluation process of cloud computing method, for the business, is associated with its execution time, the performance of the running time delay and other related indicators, this is the service quality of the business, for business application layer, should be measured by the user's experience. So, the most core or the speed of calculation and execution time, so I need to measure the simulation platform of cloud computing Hadoop core components to carry out the calculation of interface unit is optimized, actually the core of the optimization.

This article first introduces the importance and necessity of researching Hadoop, expatiates its important position in the era of big data. We can get the value and the significance of the study. Then we briefly introduce the situation of research and expound the composition of Hadoop, for subsequent core algorithm to the optimization of the necessary knowledge preparation, in the third part of this article focused on the proposed data mining algorithm, and through this algorithm, we can get more satisfactory and accurate fitting and evaluation. The final results of the model by computer got the advantages of the proposed solution, and to summarize its characteristics. Finally summarizes the whole content of the article, and puts forward some ideas for follow-up study.

## 2. Related Researches

### 2.1. The Components and Development of Hadoop

Hadoop is a kind of analysis and dealing with cloud computing and software platform for the assessment of parallel distributed computing, it has high extendibility, high efficiency, high simulation, and the advantages of reasonable structure, based on the topology of the network, according to the communication link of on-off situation automatic data retransmission, to copy the relevant data, and other functions. Can also according to the business of different data processing for the distribution of the distributed computing and parallel, which can be in the business scope of utmost to mobilize resources disposal, so that we can improve the speed of operation. Because of its open source software platform, and become the widely used large data cloud computing platform.

Hadoop data processing has high efficiency, because it has efficient data storage and efficient data processing. In efficient storage and efficient data calculation, will be a number of big data segmentation, and divided with the method of data mapping, map to each virtual node, the storage and computing unit can be used in a master-slave works, which is corresponding to the network topology structure. In storage system, the adoption of distributed storage, it is advantageous to the scheduling and transmission of data, it is good for its data decomposition, which is not affected by the processing unit capacity constraints. And it can make each storage unit's structure is simple. For each storage node, can be divided into two function block, the first function block has the scheduling and storage capabilities, and another function block is used to implement the storage nodes. For the process of data storage unit obtain data is shown in Figure1.

For cell, adopt multithreading approach, which can be through projection data and the data extraction, among them, for data clustering is a very important process of cell, are also a necessary part of the process, because of its decision behind data calculation results, but also will determine the speed of computing. For cell processing process, can according to the following process. First of all, the input files into multiple segments, after multiple segments by distribution of each calculation unit. Each calculation subunits can be allocated to different virtual workstation, the workstation is mainly controlled by the user program in procedures, and according to the mapping address each virtual workstation for input operation, according to the input operation, written to the virtual workstation, and then clustering processing, according to the specific clustering algorithm is analyzed. As you can see in this a series of operations, the most complex is the process of clustering.

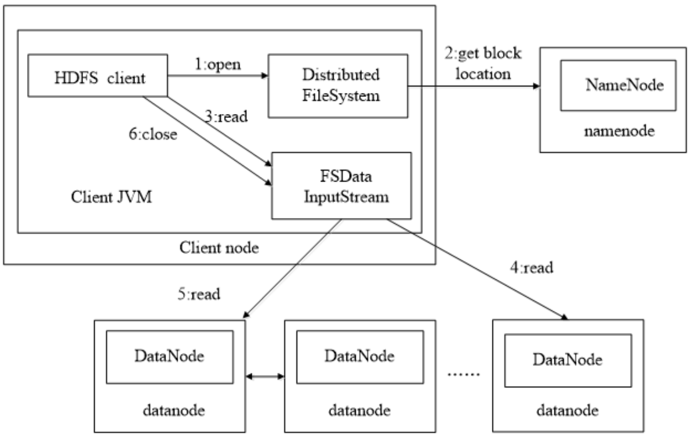


Figure 1. Data Process to be Obtained

### 2.2. The Study of the Status of Clustering Algorithm

For clustering method, there are many ways, for a variety of methods, the method of bionics is the focus of the recent research, such as: method of neural network, genetic algorithm, particle swarm optimization algorithm, ant colony algorithm and swarm algorithm, these algorithms are based on complex network, however, these networks, each link transmission of data is relatively small, so the simple traditional clustering method was applied to large complex data network, obviously its performance is not good, so according to big data complex networks is analyzed, according to the characteristics of the network, find the features of clustering method.

Here are aimed at the research status of various clustering methods were analyzed, and the first of all, the first to see to the problem of clustering optimization problem, namely is usually some search method, however, for the most optimal search method is certainly poor search method, however, this method requires a lot of computing resources, these resources through the scheduling way, will generate a lot of time delay, the delay will directly affect the performance of the system. So how to find a good, low computational complexity of the search algorithm became the core of the clustering algorithm to solve problems. Obviously two parts in iterative search algorithm performance and the complexity is to determine the size of the key core issues, namely, the choice of initial value problems and iteration convergence problem. First, determine the initial value problem, according to the location of the initial value can determine the starting point of the search. In the clustering method, the heuristic algorithm is a lot of more phyletic, including annealing algorithm, chaos algorithm and a series of methods.

And particle swarm optimization algorithm is similar to the swarm algorithm and the ant colony algorithm, which is to simulate the birds, is engaged in the development of foraging behavior and handling, and it can effectively solve the problem of the search, but can't solve the problem of complex topology structure. It is a kind of mass incidents based on a small scale, this event is simple topology, the data communication is relatively small, and its initialization is random, but because of its optimal solution according to each seemingly between searches, the search of solution must be an optimal solution. But due to the optimal solution, so need a lot of calculation, and finally affect performance. So how to compromise between complexity and the optimal solution, this would require the application of complex network transmission and exchange of data large, also is the era of big data. So to improve the particle swarm, integrated the advantages of other algorithms. Now mainly from two aspects of optimization of particle swarm optimization, on the one hand is to optimize the algorithm convergence speed, on the other hand, focuses on the specific task is optimized. But miss a key, the key is to optimize the initial value. That is the initial value of the mechanism is established.

In the particle swarm optimization algorithm, how to accelerate its convergence speed? Most of the measures are for the inertia factor and related factors of learning. These factors mainly determine the scope of its search and iteration process is also the key factors of the gradient. So the convergence speed is determined by the optimization factor. Another is the way to the search. In the case of search, global search and local search in two ways, these two approaches are to determine the size of the search scope.

For genetic algorithm, the search range is relatively small, but they tend to fall into local optimum. And according to the laws of nature, and its convergence speed is relatively slow, but relatively optimal initialization mechanism. Can be initialization mechanism of genetic algorithm can draw lessons from in particle swarm optimization algorithm to improve, to be able to get better search result, but how to make use of the complexity of the topological structure and the characteristics of huge data transmission, which can get the optimal initial value, at the same time also can get the optimization of the initial data.

In the process of search and iteration approach also varied, but these iterative algorithm, such as the gradient method, annealing algorithm, such as a variety of methods, but all have the ability to search performance, in standard particle swarm optimization algorithm, the main suitable for continuously differentiable function to search, but the values of discrete function, its performance will become very poor. At the same time, for the good of improved particle swarm optimization (pso) how to apply the mathematical theory, to evaluate the various aspects of performance, thus a more scientific was improved, it is also a need to study now. And on the scope of application and characteristics is an important research content of need. Implemented at the same time need to use what kind of form is an important exploration made by this paper. The following summarizes the shortcoming of particle swarm, and can improve the measures are put forward.

First, rely mainly on the topology of particle swarm optimization (pso) node is particle interaction to achieve the search, but its own does not change the corresponding mechanism, so easy to make the whole search get into local optimal solution, search the main fixed phase change. In addition, the second, in the selection process of particle swarm optimization, due to its termination condition is too harsh, can't stop, so will cause the algorithm be death cycle. In either case, the particles interact, mutual restrict, make the unnecessary consumption, leading to unable to convergence, thus increasing the complexity of the algorithm.

## 3. The Proposed Scheme

### 3.1 Initialization Model

For the initialize model, according to the law of large numbers of random probability, in this paper we consider the normal distribution model, for different data can be composed of mixed normal distribution. This kind of mixed normal distribution can be obtained by Taylor series expansion way to approximate any function, but also through this mechanism, can be suitable for the majority of cases, especially in complex networks, and process large data transmission in among the topologies, assumed in the mixed normal distribution, large data can be used to $D$ , which contains the dimensions $N$ of the sample, and we can obtain the probability density function of the $k$ data, it can be expressed as (1)

$$p(x,\beta) = \sum_{i=1}^{k} \alpha_j N\left(x,\eta_j,\Xi_j\right) \tag{1}$$

In the formula (1), function $N\left(x,\eta_j,\Xi_j\right)$ can be represented as

$$N\left(x,\eta_j,\Xi_j\right) = \frac{\exp\left[-\frac{1}{2}\left(x-\eta_j\right)^T \Xi_j \left(x-\eta_j\right)\right]}{\left(2\pi\right)^{d/2}\left|\Xi_j\right|^{1/2}} \tag{2}$$

According to this dimension normal distribution can be combined, according to the different dimensions of random data to establish the initial value, but the scheme of seemingly random, it is actually a pseudo random process, which can be adjusted according to the initial data automatic value, source is also random adjustment search. The specific process, in this model, first, not the initial value assigned to a fixed location, but according to the data and data classification situation (*i.e.* samples) to decide the final process.

Step 1: determine the initial of a discrete data set;

Step 2: according to the normal distribution parameters of initial set to top the initial, that is $\left(x,\eta_j,\Xi_j\right)$

Step 3: according to the normal distribution, we can get the initial node;

Step 4: according to the initial node, can be used to measure the iterative algorithm, with the mathematical mean value;

Step 5: if the expectation value is higher than a fixed sample expectation, you can think of this distribution is close to the global optimal solution that is the starting point of the search.

The particle swarm optimization algorithm, according to the test function in a population of particles, in order to test this method of random initialization values, the random initialization of testing that is the calculation of normal multivariate distribution of these values and the test function of the minimum Euclidean distance, the specific relationship as shown in formula (3),

$$d = \sqrt{\left[f(x,\beta) - p(x,\beta)\right]^2} \tag{3}$$

In the formula (3), the test function can make many kinds of function, one of the most science is

$$f(x,\beta) = \sum_{i=1}^{k}\left[x_i^2 - 10\cos\left(2\pi x_i\right) + 10\right] \tag{4}$$

### 3.2. The Plan of Standard Particle Swarm Optimization

For standard particle swarm plan, need to consider from the following aspects, first, is that each particle swarm elements, known as particles, also is the optimal solution of the every possible scope, for the second is a collection of particles, which is the combination of the approximate optimal solution. The third part is within the scope of each particle in the search space, and the fourth part is to search the standard, need to determine the iterative search criteria. The fifth part is about the iteration results after the degree of match and the optimal solution. In $k$ dimension of search space and each particle can be expressed as a $k$ dimensional space, each space location can also use this.

According to the definition of all, we can know each scope of the particle swarm method; they can be obtained by the following two formulas, namely

$$L(n+1) = L(n) + \lambda_1 m_1 \left[\theta(n) - x(n)\right]$$
$$+\lambda_2 m_2 \left[\theta(n) - x(n)\right] \quad (5)$$

Where $\lambda_1$ and $\lambda_2$ represent each particle and each particle to go forward in the direction of the forward speed, and $L(n)$ is the location of each particle at last time.

$$\theta(n+1) = \theta(n) + x(n+1) \quad (6)$$

From it which can be seen that the particle swarm method, decision factors are $\lambda_1$ and $\lambda_2$, each function set also need to consider, which determines the speed of convergence.

According to the method of particle swarm, we can get its execution process; this process can be divided into the following steps:

1. The parameters configuration, that is, setting each parameter, $\lambda_1$, $\lambda_2$, $\theta(n)$ and $x(n)$;
2. Determination criterion for each iteration;
3. According to the particles and iterative standard, we can get the iterative pick integral function, namely (5);
4. The use of each iteration, namely (5) and (6) to determine the location of the particle and the specific search range;
5. Checking whether achieve the optimal or not;
6. If the optimal, stopping the iteration, determining the optimal solution.

### 3.3. The Improved Particle Swarm Algorithm

In improved particle swarm optimization algorithm, the first to say that in Section 3.1 of initialization is also included in the improved particle swarm algorithm, in this part, the mainly is to consider is the iterative parameter settings, and other iterative formula, so this part is the core part of the improved particle swarm algorithm, in Section 3.4 describes how to implement in the Hadoop, program realization method.

In this article, based on genetic algorithm, using a deviating from the comprehensive matrix, alignment set, this can lead to faster particles to the near optimal solution, so the main to find deviation matrix, the matrix to determine the specific direction of particle movement, to the final solution of traditional particle swarm optimization algorithm into local optimum situations. Assuming the global deviation matrix for (7)

$$C = \begin{bmatrix} 0 & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & 0 \end{bmatrix} \tag{7}$$

In this deviation matrix, each element can be value for the three values, these three values are discrete values, according to the discrete values can gradually become search method, it is +1,-1 and 0. The direction of the deviation of the particle can be sure, because it is discrete values, they can be better in the program implementation.

According to deviation matrix, we can find the improvement of (5), the improvement is one of the parameters, it can be improved in the form of (8),

$$L(n+1) = L(n) + \lambda_1 m_1 \left[\theta(n) - Cx(n)\right]$$

$$+\lambda_2 m_2 \left[\theta(n) - Cx(n)\right] \tag{8}$$

According to (8), of course, it also needs to accelerate the relevant parameters were optimized, such as $\lambda_1$ and $\lambda_2$, $\theta(n)$, $x(n)$, they may represent the following equation , such as equation ( 9-12 )

$$\lambda_1 = \frac{\left\|(C)_i\right\|_\infty^2 - \left\|(C)_j\right\|_\infty^2}{\left\|(C)_j\right\|_F^2} \tag{9}$$

$$\lambda_2 = \frac{\left\|(T)_i\right\|_\infty^2 - \left\|(T)_j\right\|_\infty^2}{\left\|(T)_j\right\|_F^2} \tag{10}$$

$$\theta(\mathrm{n}) = \frac{\left\|(TT)_i\right\|_2^2 - \left\|(TT)_j\right\|_2^2}{\left\|(TT)_j\right\|_F^2} \tag{11}$$

$$x(n) = \frac{\left\|(TT)_i\right\|_\infty^2 - \left\|(TT)_j\right\|_\infty^2}{\left\|(TT)_j\right\|_F^2} \tag{12}$$

Therefore, we can obtain the process of the improved particle swarm optimization algorithm:

1. The configuration parameters, which set each parameter, $\lambda_1$, $\lambda_2$, $\theta(n)$ and $x(n)$ ,see formula(9-12);
2. Determine for each iteration standard;
3. According to the particles and iterative standard, we can get the iterative pick integral function, namely formula (8);
4. Using each iteration, namely formula (8) and (6) to determine the location of the particle and the specific search range;
5. Checking whether achieve the optimal or not;
6. If the optimal, stopping the iteration, determining the optimal solution.

### 3.4. The Software Implementation of Hadoop

Specific algorithm based on the above, this paper mainly with the method of frame to elaborates the specific implementation, in which mainly to read data from the global, the core algorithm is relatively simple to implement. According to the process is through the fragments of storage, all of the data samples, through sampling, determine initialization searching initial point. The second step, according to the read data using a relational database, search, here at the core of the search algorithm is proposed in this paper improved particle swarm algorithm here need to emphasize that is initialized, the normal distribution of initialization is done in the first step.

## 4. Simulation Results

The simulation by the developed on the basis of distributed parallel computing, the use of the resources of the platform can be through the computer simulation data storage and transmission, but its lack of topology change link, according to the reality, this experiment is modified, the modified according to the topology changes its transmission of data and transmission time. The simulation environment, include computer configuration environment. Computer configuration environment is the operating system is Windows 7, memory is 8 g, hard disk capacity is 1 t. Level of the simulation amount of data in millions, under the condition of millions of level, it can be seen that due to the initialization of the rationality of the design, there have been big reduce the execution time of the data in the Hadoop, as shown in Table 1.

**Table 1. The Data Execution Time of MapReduce in the Hadoop**

| Map Task Number | Record number | Data Size | Original Method Excution time(sec.) | Proposed Method Excution time(sec.) |
|---|---|---|---|---|
| 10 | 1000000 | 100M | 33 | 31 |
| | 5000000 | 500M | 51 | 48 |
| | 10000000 | 1.0G | 98 | 80 |
| | 50000000 | 5.0G | 228 | 197 |
| 40 | 1000000 | 100M | 29 | 25 |
| | 5000000 | 500M | 32 | 33 |
| | 10000000 | 1.0G | 40 | 34 |
| | 50000000 | 5.0G | 146 | 123 |
| 80 | 1000000 | 100M | 30 | 23 |
| | 5000000 | 500M | 45 | 32 |
| | 10000000 | 1.0G | 51 | 26 |
| | 50000000 | 5.0G | 178 | 95 |

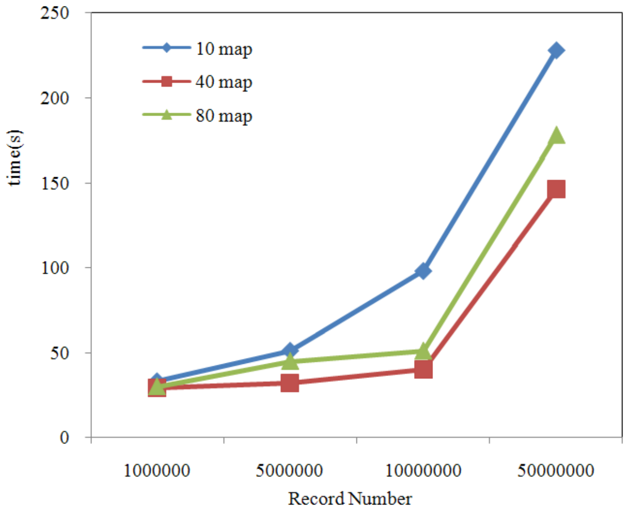


**Figure 2. Original Method Excution Time**

According to the simulation can be seen in Figure 2, as the record number is different, its execution time also will increase, but in the same record number, the number of map, the more the number of its data processing ability also is stronger, its execution time is less, so increasing the number of map is worth. It is one of the effective methods to improve the execution efficiency. But when the number of Map more than a certain amount of time, it will reduce the operational efficiency of execution. This is because with the increase number of map, the Hadoop performance problems gradually exposed, so the job execution efficiency reduced. It can be improved by using the particle swarm optimization algorithm mentioned above.

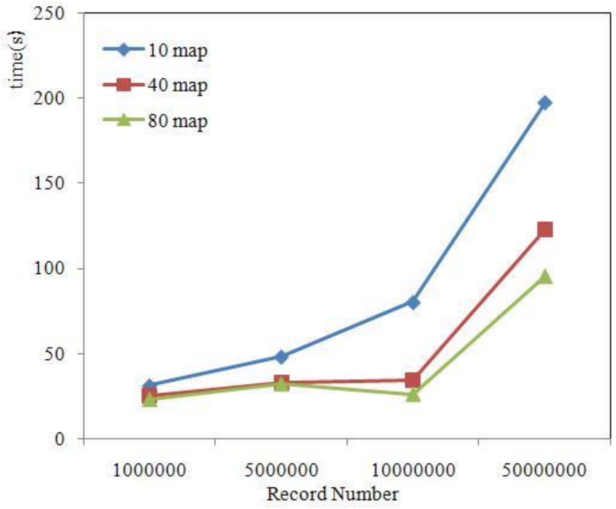


**Figure 3. Proposed Method Excution Time**

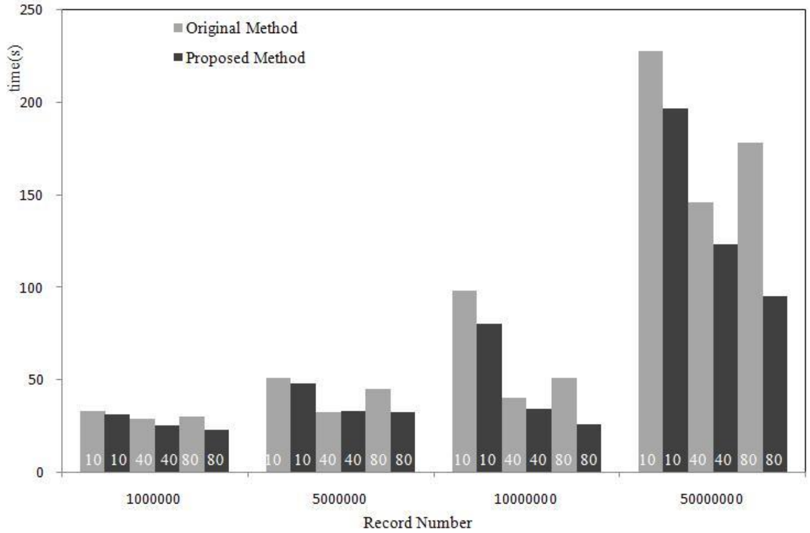


**Figure 4. Comparison Chart**

In Figure 3, when number of map is much, the implementation of the database into execution time also is less. But it can be seen in this paper, the proposed algorithm, for different data, namely, big data processing has very strong robustness, the time also is relatively small, improved the processing speed. We can find that in Figure 4, for using the proposed algorithm, execution time is less compared with using original method in the Hadoop. This proves that the effectiveness of the improvement in the Hadoop. So the proposed algorithm is validated by the simulation can be advanced, and it can be used in the Hadoop platform in the future.

## 5. Conclusions

In this paper, according to the complex interconnection network characteristics of the big data, the number of the transmission and communication of data is large, and its matching search efficiency is lower. It uses the better structure of the particle swarm algorithm and initializes, as well as the search direction problem put forward different solutions. Through simulation comparison, the proposed scheme has better performance and shorter execution time. It provides beneficial experience for the subsequent improvement of Hadoop platform.

## References

[1] Dinh H. T., Lee C. and Niyato D., "A survey of mobile cloud computing: architecture, applications, and approaches", Wireless communications and mobile computing, vol. 13, no. 18, **(2013)**, pp. 1587-1611.

[2] Garg S. K., Versteeg S. and Buyya R., "A framework for ranking of cloud computing services", Future Generation Computer Systems, vol. 29, no. 4, **(2013)**, pp. 1012-1023.

[3] Iosup A. and Epema D., "On the Gamification of a Graduate Course on Cloud Computing", The International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, **(2013)**.

[4] Venkata K. P., "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing, vol. 13, no. 5, **(2013)**, pp. 2292-2303.

[5] Ryan M. D., "Cloud computing security: The scientific challenge, and a survey of solutions", Journal of Systems and Software, vol. 86, no. 9, **(2013)**, pp. 2263-2268.

[6] Szymanski T. H., "Low latency energy efficient communications in global-scale cloud computing systems", Proceedings of the 2013 workshop on Energy efficient high performance parallel and

distributed computing. ACM, **(2013)**, pp. 13-22.

[7] Amoda N. and Kulkarni R. K., "Efficient Image Retrieval using Region Based Image Retrieval", International Journal of Applied Information Systems (IJAIS)–ISSN, **(2013)**, pp. 2249-0868.

[8] Pérez O., Amaya I. and Correa R., "Numerical solution of certain exponential and non-linear Diophantine systems of equations by using a discrete particle swarm optimization algorithm", Applied Mathematics and Computation, vol. 225, **(2013)**, pp. 737-746.

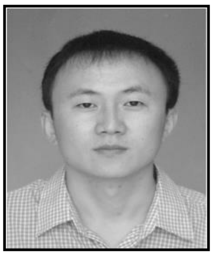
[9] Mandal D., Kar R. and Ghoshal S. P., "Digital FIR filter design using fitness based hybrid adaptive differential evolution with particle swarm optimization", Natural Computing, vol. 13, no. 1, **(2014)**, pp. 55-64.

[10] Belmecheri F., Prins C. and Yalaoui F., "Particle swarm optimization algorithm for a vehicle routing problem with heterogeneous fleet, mixed backhauls, and time windows", Journal of intelligent manufacturing, vol. 24, no. 4, **(2013)**, pp. 775-789.

[11] Katherasan D., Elias J. V. and Sathiya P., "Simulation and parameter optimization of flux cored arc welding using artificial neural network and particle swarm optimization algorithm", Journal of Intelligent Manufacturing, vol. 25, no. 1, **(2014)**, pp. 67-76.

[12] Zhang L., Chen Q. and Miao K., "A Compatible LZMA ORC-Based Optimization for High Performance Big Data Load", Big Data (BigData Congress), 2014 IEEE International Congress on. IEEE, **(2014)**, pp. 80-87.

## Author



**Jia Min-zheng**, received the Master's degree in Technology of Computer Application from China University of Mining & Technology Beijing in 2006. He is currently working in Beijing Polytechnic College. He is currently researching on Network Measurement, Big Data and Cloud Computing. Beijing Polytechnic College Scientific Research Task (bgzykyz201405).