Research on Cloud Resource Hadoop Cluster Scheduling Algorithm Based on Artificial Bee Colony Algorithm

Fan Haiyan

Department of Information, Jiangsu College of Finance & Accounting, Lianyungang, Jiangsu, 222061, China

email: 93237483@qq.com

Keywords: Bee Colony Algorithm, Cloud Resources, Hadoop Cluster Scheduling Algorithm

Abstract: Today, the number of data and users enterprises are facing is increasing. At the same time, cross industry access to business systems (traffic, data storage and management pressure are higher than in the past). These "big users, big data, big systems" have become problems. The necessary conditions for the development of cloud computing. With the development of information technology, cloud computing has become an important solution to the challenges of information and social big data. Cloud computing applications put forward higher and higher requirements for the quality of service of virtual resources. Is cloud task scheduling a key problem affecting the quality of service of virtual resources? It has become a research hotspot. Swarm intelligence algorithm is an important analysis method to solve optimization problems. In recent years, it has achieved good results in scheduling problems.

1. Introduction

With the wide application of computer and network communication technology, the rise of the special Internet, the gradual maturity of the Internet, the emergence of mobile Internet, cloud computing, the Internet of things, social networks and other next-generation technologies, the sources of data generation have multiplied. Data is also growing exponentially, and the characteristics of traditional data are changing. Therefore, today's society has brought a "big data era"[1]. Big data has the characteristics of large amount of data, many types, fast data flow and great potential value. It makes sense to process and use big data. Don't open up cloud computing. Since the concept of cloud computing was put forward, it has been widely concerned by various companies, research institutions and governments. Both architecture and key technologies have made great progress. Cloud computing has completely changed the IT architecture and it business model. The so-called relationship between cloud computing and big data can be said that the development of cloud computing technology and architecture has accelerated the society into the "big data era", and the change of big data has promoted the development of cloud computing technology and related applications [2]. For example, Hadoop is the most popular cloud computing distributed storage computing platform, and it is the key mainstream tool of big data storage processing and query analysis technology.

2. Significance of Research

Old m, Microsoft, HP and other overseas enterprises have always affirmed the importance of China's cloud computing market, and have successively launched various cloud computing products and solutions. During the period of listing, it is an important research topic to improve customer satisfaction and save the resources of service providers. In order to improve the use of cloud computing, there are various methods and algorithms [3]. This paper focuses on two important aspects: cloud task scheduling and suwoo intelligent algorithm. With the large-scale production of data, cloud computing data processing technology itself has been constantly improved, and its service objectives have been expanding. After several stages of the development of the artificial bee colony algorithm, the research on the artificial bee colony algorithm has made some achievements

in scheduling application. However, in the scheduling problem, the basic discrete artificial bee colony algorithm is often used after the simple discrete processing, and the solution can not be compared with the classical scheduling algorithm. Most of the existing Jinyun task scheduling models are also considered in a single way. They do not have these two problems and cannot meet the requirements of users and service providers. Therefore, it is of great significance for the development of cloud computing users and suppliers to build a comprehensive cloud task scheduling model and obtain a comprehensive solution with high satisfaction.
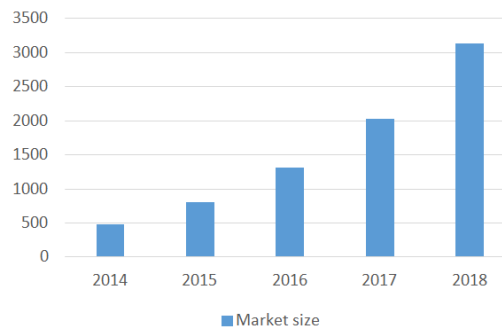

Figure 1 Cloud computing market scale

3. Analysis of the Current Situation of Bee Colony Algorithm

The swarm algorithm is similar to other soom intelligent optimization algorithms [4]. It is a new found algorithm of biological warmth optimization by simulating the social behavior of natural population. The biological mechanism of pupae is simulated by bee and ant colony algorithm, which can be divided into two types according to different actions: one is triggered by bee's reproductive behavior mechanism. These two heuristic algorithms have their own unique simulation mechanism and development history. As for the swarm algorithm based on the behavior of bee reproduction, Abass first proposed the optimization model of bee reproduction, M 0. After that, bozorg Haddad and a. afshar improved the model and applied it to the import of discrete variables. Reservoir optimization problem W. different types of test platform m for LAN math problems. However, the so-called bee colony algorithm mainly refers to the bee colony algorithm which simulates the bee's baiting action. There are many researches on the bee colony algorithm based on baiting action. In the study of pupae based on goat's feeding behavior, poplar first proposed a virtual bee algorithm using puppets to solve the optimization problem. Then, in order to solve the problem, other classical algorithms (differential evolution, particle swarm optimization, etc.) are compared to prove the effectiveness of the proposed algorithm. Artificial swarm is a dimension of algorithm today[5]. At present, the artificial bee colony algorithm proposed by the research team is a relatively systematic bee colony intelligent biological algorithm proposed by Wang Erqi scholar karaboga in 2005. w. Function optimization training neural network W, data mining, in other fields, it makes good use of the algorithm. Discrete artificial clustering algorithm (now widely used in scheduling problem research).
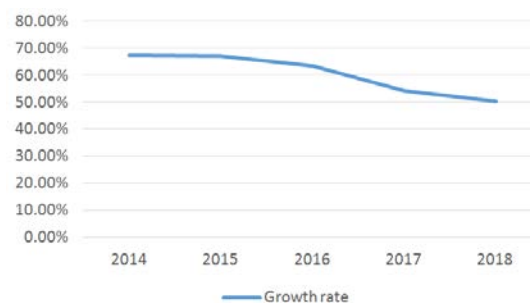

Figure 2 Growth rate of cloud computing market scale

4. Artificial Bee Colony Algorithm

4.1 Basic Theory of Artificial Bee Colony Algorithm

Artificial bee colony algorithm is one of the most widely used and simple intelligent optimization algorithms. Artificial bee colony algorithm has experienced a development period, which is the most mature result of bee colony algorithm, and has a very extensive application value.

4.2 Principle of Artificial Bee Colony Algorithm

Artificial bee colony algorithm is a bionic swarm intelligence algorithm proposed by calaborga, the main researcher of artificial bee colony algorithm in 2005. This is a kind of intelligent behavior that imitates the emergence of bee free mechanism in nature [6]. Using the transformation of honeybee and the function of honeybee, the optimal location of honey source is realized. A kind of

4.3 Characteristics of Artificial Bee Colony Algorithm

The process of foraging and searching for food in bee colony algorithm, the reflection of four characteristics in system learning, integrity, order, correlation and dynamics. Holism means that the whole colony shows an action that an individual can't show, and the overall ability of bee ants is far greater than the sum of individual abilities; orderness means that they work in a neat way according to their own rules, and are not disturbed by external information. Then, each iteration of the swarm algorithm is updated by dynamic action, and various information such as the current best solution of the algorithm is changed [7]. The honey source will be collected, and the position of the honey source will appear on the artificial bee. When it is difficult to solve the problem by decentralized action, usually, the first few small and simple segmentation problems of these large-scale difficult problems, and these small and simple parallel segmentation problems, can be dealt with. Finally, the final answer to the first question is obtained. Besides simplification, distributed has other advantages. Even if one unit in the system cannot process the result, it will not affect the whole performance. Artificial swarm algorithm is a kind of decentralized idea, which is assigned to each bee to find the best solution. According to their own inherent actions, it can search the executable solution of bee source independently. In summary, the algorithm moves in a better direction. In this process, some individual bees can not find a good solution, or can not affect the overall algorithm, so as to find the best solution to solve the problem of deviation from the established orbit.

4.4 Steps of Artificial Bee Colony Algorithm

In ABC algorithm, there are h kinds of bee splitting: bees and astrolabes [8]. A sheep bee that keeps itself looking for honey. Scotby randomly looks for honey sources in the solution area, and follows bees in the information sharing area, relying on bees to transport honey source information, so as to select the collected honey sources. In the artificial bee colony algorithm, the location of the honey source represents the executable solution of the optimization problem, and the abundance of the honey source represents the quality of the executable solution. ABC algorithm initially initializes n executor sources in d-Dimension solution space. Each honey source can lure baboons, so n honey sources in the space can lure n sheep adopted by W mother group. The position of the bee is the position of the pupa. In the process of algorithm, the number of honey sources will not change with the repetition process. Hire helper bees in the dance area because bees and honey bees share the information being transported in order to attract followers in proportion to the quality of their honey source. The next bee then selects the operator, which is set by the algorithm to decide which honey source to go to. After each honey source arrives, the follow-up search is performed at the honey source, and the position of the newly obtained honey source is compared with the original work. Compare the honey source corresponding to honeybee. If the new nectar is better than the original nectar source, the new nectar will be used as the new nectar source and collected by the employed bees; otherwise, the original nectar source will continue to be mined and the corresponding amount of nectar source will be collected. All the honey has attracted employed bees, preparing for a new round of mining development. If the same sweet source is mined several times or more, the quality of the solution has not been improved. At this time, the beekeeper who collects

the honey source will become the astrolabe, enter the astrolabe ratio stage, and then the astrolabe will enter the solution space [9]. Machines generate new solutions to replace the original Hanni source.

Table 1 Comparison of intelligent optimization algorithms

Algorithm name	Describe	Advantage	Shortcoming
Genetic algorithm	Simulation of reproduction in natural selection and natural inheritance	Large scale global search capability	Fail to make good use of system feedback
Differential evolution algorithm	Heuristic based on group difference	Simple principle	Using difference strategy in mutation operation
Ant colony algorithm	Information transmission between ant groups	With positive feedback mechanism	Lack of pheromone in the early stage

5. Parameter Analysis of Artificial Bee Colony Algorithm

In the artificial bee colony algorithm, the basic algorithm stage has been completed by searching the bee hive. Behavior employs bees to carry out local mining near the honey source, including using bees to deliver the corresponding honey source quality information, and calculating the resilience used by bees.

6. Conclusion

Under the background of big data explosion, the effective data processing method with cloud computing faces the dual challenges of efficiency and effect [10]. How to deal with information quickly and accurately in cloud computing, in order to meet the needs of customers, it is the most important research project for us to think about the load balance of cloud computing resources.

References

[1] YoungJu Moon, HeonChang Yu, Joon-Min Gil,. A slave ants based ant colony optimization algorithm for task scheduling in cloud computing environments. Human-centric Computing and Information Sciences, vol. 7, no. 1, pp. 28, 2017.

[2] Xuan Chen, Dan Long. Task scheduling of cloud computing using integrated particle swarm algorithm and ant colony algorithm. Cluster Computing, no. 4, pp. 1-9, 2017.

[3] CHEN Xuan, XU Jianwei, LONG Dan. Resource scheduling algorithm of cloud computing based on ant colony optimization-shuffled frog leading algorithm. Journal of Computer Applications, 2018.

[4] Soualhia M, Khomh F, Tahar S. A Dynamic and Failure-aware Task Scheduling Framework for Hadoop, no. 99, pp. 1-1, 2018.

[5] Komarasamy D, Muthuswamy V. Content-Based Federated Job Scheduling Algorithm in Cloud Computing, vol. 10, no. 2, pp. 52-59, 2017.

[6] Jeyalakshmi, S, Ram, N. Sankar. Cloud Scheduling Algorithm for Parallel Jobs Using Priority Queue. Journal of Computational & Theoretical Nanoscience, vol. 14, no. 7, pp. 3120-3124, 2017.

[7] Peiquan Jin, Xingjun Hao, Xiaoliang Wang,. Energy-Efficient Task Scheduling for CPU-Intensive Streaming Jobs on Hadoop. IEEE Transactions on Parallel and Distributed Systems, no. 99, pp. 1-1, 2018.

[8] Z. Deng, M. Fu, Q. Zhang. A research of time slot optimization task scheduling strategy based on the hadoop. Journal of Northwestern Polytechnical University, vol. 35, no. 1, pp. 32-37, 2017.

[9] Nguyen Duy Dat, Vo Ngoc Phu, Vo Thi Ngoc Tran,. STING Algorithm Used English Sentiment Classification in a Parallel Environment. International Journal of Pattern Recognition & Artificial Intelligence, vol. 31, no. 7, 2017.

[10] Yildirim G, Hallac İ R, Aydin G, et al. Running genetic algorithms on Hadoop for solving high dimensional optimization problems, 2018.