Budget constrained eddy current search intelligent optimal scheduling algorithm for Cloud Computing

Shuguang Cui

Information Engineering Departement, Hunan Automotive Engineering Vocational College, Zhuzhou, Hunan, 412001, China

Keywords: Cloud computing, eddy current search, scheduling algorithm

Abstract: In order to deal with the complex and changeable cloud computing environment, the research of scheduling problem is full of challenges, and it is very important to grasp the scheduling direction in cloud computing environment. Therefore, based on the research of traditional parallel distributed scheduling, this paper analyzes the essence of scheduling problem, and then points out the development trend of scheduling algorithm in cloud computing environment combined with eddy current search algorithm

1. Introduction

Cloud computing is based on the increase, use and delivery mode of Internet related services. It provides dynamic, scalable and often virtualized resources through the Internet. Its core uses the network to connect a large number of idle computing resources for unified scheduling and management, forming a very powerful resource pool to provide users with required services on demand. Hybrid heterogeneous cloud computing refers to cloud computing services that mix private cloud and public cloud. It can integrate the advantages of private cloud computing and public cloud computing services, achieve good coordination between them, and bring the best application experience integrating the two to enterprise users. It is the latest development trend of cloud computing.

Cloud computing is the dream pursued by human computing. The realization of the dream depends on an efficient and reliable cloud resource management system, that is, the construction and management of cloud platform is the key. Resource management is the core content of cloud computing. The purpose is to shield the heterogeneity and complexity of underlying resources and manage a variety of Scattered Cloud computing resources, so as to achieve effective resource sharing and collaborative work, Improve resource utilization. Allocate reasonable cloud computing resources for user tasks on demand to achieve load balance. As the most important means of resource management system, scheduling affects the effect of cloud resource management. Scientific scheduling can not only greatly improve the resource utilization of cloud environment, but also fully develop the heterogeneous characteristics of resources and highlight the resource personality, Make the best use of everything. Moreover, through scheduling, it can also effectively deal with emergencies such as resource entry and exit, node failure and resource failure in the cloud environment, realize accident tolerance, improve system stability and ensure service quality.

2. Scheduling problem analysis

Cloud computing is the dream pursued by human computing. The realization of the dream depends on an efficient and reliable cloud resource management system, that is, the construction and management of cloud platform is the key. Resource management is the core content of cloud computing. The purpose is to shield the heterogeneity and complexity of underlying resources and manage a variety of Scattered Cloud computing resources, so as to achieve effective resource sharing and collaborative work, Improve resource utilization. Allocate reasonable cloud computing resources for user tasks on demand to achieve load balance. As the most important means of resource management system, scheduling affects the effect of cloud resource management.

Scientific scheduling can not only greatly improve the resource utilization of cloud environment, but also fully develop the heterogeneous characteristics of resources and highlight the resource personality, Make the best use of everything. Moreover, through scheduling, it can also effectively deal with emergencies such as resource entry and exit, node failure and resource failure in the cloud environment, realize accident tolerance, improve system stability and ensure service quality.

3. Five elements of dispatching

3.1. Calculation task T

It refers to computer programs, processes, jobs or services that meet the needs and functions of users. The needs of users vary greatly, and there are various types of computing tasks. According to the coupling degree of tasks, they can be divided into independent tasks and divisible tasks. According to the correlation between tasks, they can be divided into meta tasks and dependent tasks, Among them, dependent tasks are usually represented by DAG diagram. On the basis of considering the correlation and the factor of task scale, tasks can be divided into meta task pool and workflow. According to the intensity characteristics of task resource requirements, tasks can be divided into computing intensive, data intensive, communication intensive and I / O intensive tasks. According to the nature of services, tasks can be divided into scientific computing, transaction processing Real time processing tasks can be divided into uniform, periodic, random and batch tasks according to the arrival law. According to the priority, it can be considered as preemptive, non preemptive tasks, etc.[2].

3.2. Implement resource and environment R

It refers to the running platform of computing tasks, which is composed of processor, memory, I / O equipment, network and other hardware resources and software support systems. According to the computing mode, it can be divided into single machine system and multi machine system. Multi machine system can also be divided into parallel computing, distributed computing, cluster computing, grid computing, mobile computing, pervasive computing, P2P computing Cloud computing system, etc. according to the architecture, parallel computing system can be divided into SIMD, MIMD, PVP, SMP, MPP, cow, DSM, etc. according to the type and nature, it can be divided into homogeneous system and heterogeneous system. The cloud computing environment to be studied in this paper belongs to heterogeneous and distributed multi machine system.

3.3. Dispatching objective G

It is the highest, or specific, or focus requirement in the scheduling process. There are many scheduling objectives according to different needs, different execution environments and different occasions. The scheduling objectives from users include: shortest completion time, deadline, service cost, correctness, fairness, security Reliability, etc. the scheduling objectives from the server or resource manufacturer include: resource utilization, revenue, operation cost, throughput, fault tolerance, load balancing, reputation, energy consumption, etc. the scheduling objectives from the scheduler itself include: response speed, efficiency, success rate, steady state, intelligence, etc.

3.4. Scheduling constraint C

It is a general requirement in the scheduling process, that is, the restrictive conditions that must be met, which is the restriction on the rationality of scheduling decision. Scheduling constraints are expressed as a conditional relationship, which limits the excessive pursuit of specific scheduling objectives to ensure the reasonable implementation of scheduling. There is a complex relationship between scheduling objectives and constraints. When the intensity or level of requirements changes, the objectives may become constraints, For example, when the scheduling target T is the minimum completion time, the cost and resource scale may be scheduling constraint C. However, if the scheduling target is the minimum cost, the completion time and power consumption may become scheduling constraint C.

3.5. Scheduling policy P

It focuses on describing the specific implementation scheme of scheduling, which is the ideological basis for the formation of scheduling algorithm. According to the timing of scheduling policy decision, scheduling can be divided into static and dynamic scheduling. According to the considered system scale, scheduling can be divided into local and global scheduling. According to the different executors of scheduling, Scheduling can be divided into other scheduling and self scheduling. Scheduling can be divided into adaptive scheduling and non adaptive scheduling according to whether the scheduling strategy is automatically adjusted with the change of system state information. Scheduling can be divided into preemptive scheduling and non preemptive scheduling according to whether tasks are allowed to be migrated and executed when the scheduler performs scheduling operations, Scheduling can be divided into migration and non migration scheduling. Considering where the scheduling algorithm is executed, where the scheduling information is stored, and how complex the technology used by the scheduling algorithm is, the scheduling methods can be divided into central scheduling, hierarchical scheduling and distributed scheduling. In addition, the scheduling strategy also includes heuristic methods, which can be divided into table scheduling, cluster scheduling Replication scheduling and non deterministic scheduling [3].

4. Eddy current search algorithm

Vortex search (VS) is a Turkish scholar Berat dog An and tamer lmez proposed a novel heuristic optimization algorithm based on random search and pattern search, which was inspired by the vortex generated when stirring liquid. The eddy current search algorithm adopts a strategy of adaptively adjusting the search radius according to the number of iterations. It has fewer parameters and rapid iteration, and can find the optimal solution in a short time.

Although the eddy current search algorithm has the above advantages, the algorithm itself has some defects, such as the single generation mode of alternative solutions and easy to fall into local optimal solutions in the later stage. In order to solve the above problems, this paper proposes a chaos vortex search (CVS) algorithm based on chaos theory, that is, the alternative solution of the eddy current search algorithm is updated by chaotic time series. Because the chaotic sequence variables generated by chaos theory have the characteristics of ergodicity, randomness and regularity, and traverse all States without repetition according to their own laws in a certain range, the current optimal alternative solution is mapped from solution space to chaotic space to generate chaotic time series, and the chaotic time series is mapped from chaotic space to solution space for iterative optimization, So as to increase the diversity of alternative solutions of the algorithm [4].The principle of eddy current search algorithm is as follows

Let the dimension of the solution space be d dimension, the value range of the j dimension be $E_j = \left\lfloor e_{lowerj}, e_{upperj} \right\rfloor, j = 1,2,...,D;$, and the number of alternative solutions be n; The maximum number of iterations is maxitr; RJ is the current search radius and I is the current number of iterations.

Taking the solution space center $u_0 (u_0 = \left[u_{01},...,u_{0j},...,u_{0D}\right], u_{0j} = \frac{e_{lowerj} + e_{upperj}}{2})$ of the objective function as the center of the circle, N random alternative solutions are generated by Gaussian distribution around the center of the circle, and the probability density function satisfies the formula:

$$p(x|\mu)^T \sum {}^{-1}(x-\mu) \exp\left\{-\frac{1}{2}(x-\mu)^T \sum {}^{-1}(x-\mu)\right\} \quad (1)$$

Before the iteration begins, σ Initial value of σ 0 is determined by the upper and lower limits of

the solution space of the objective function:

$$\sigma_0 = \frac{\max(e_{upperj}) - \min(lowerj)}{2} \tag{2}$$

With the increase of the number of iterations I, the value of AI will gradually approach 0 from 1, so that the function (1 / x) · gamma inch ν The value of (x, a) will also be gradually reduced, so as to realize the adaptive adjustment of the search radius Ri with the number of iterations I.

5. Overview of scheduling algorithms in cloud computing environment based on eddy current search algorithm

Cloud computing environment is different from the previous scheduling execution environment. At present, domestic and foreign scholars have proposed different scheduling algorithms according to the characteristics and requirements of cloud environment. Huang et al. Proposed a dynamic distributed scheduling algorithm Casa according to the characteristics of resource distribution in cloud environment. In this algorithm, each scheduling node has a meta scheduler, It is used to receive tasks submitted by local users and allocate tasks. At the same time, each scheduling node also maintains information sharing to achieve the purpose of load balancing between nodes. Lee et al. On the contradiction caused by expanding resource sharing in the cloud environment, that is, by improving resource utilization, they expand the economic benefits of cloud service providers, However, it reduces the performance level required by users, such as the contradiction that tasks cannot be completed within the completion time. A new revenue calculation model is designed. This model not only considers the benefits brought by the current service, but also considers the benefits brought by other tasks, and a new service request algorithm is proposed, Under the constraints of user completion time requirements, the goal of maximizing the interests of service providers is achieved.

In the cloud computing environment, the adopted cloud computing scheduling strategy will directly affect the execution efficiency of user tasks and the resource utilization of cloud computing system. In order to ensure the service quality of users, it is of great significance to select a better task scheduling strategy and improve the resource utilization of cloud computing system [5].

At present, there are many researches on cloud computing task scheduling and resource allocation. Because each cloud service provider's cloud platform is developed on its own infrastructure, and the structure of some infrastructure is different, there are no unified specifications and standards for many problems of cloud computing. Cloud computing evolved from the deep development of distributed computing and grid computing in terms of application scope, commerciality and technology. The research on task scheduling and resource allocation in distributed computing and grid computing has been quite mature, which provides a certain reference for task allocation in cloud computing environment.

However, there are many differences between cloud computing and grid computing, such as business model, application type, network protocol system and so on. Cloud computing will not hand over specific tasks to the actual physical node computer for execution. It uses mature virtualization technology to map the host resources into multiple virtual machines to form a virtual resource pool, and the tasks submitted by users are handed over to the virtual resource pool for execution. Cloud computing task scheduling is to organize and schedule various forms of virtual machine resource pools to execute the tasks submitted by users according to the needs of users [6].

Cloud computing aims at business services and users are the center. In the face of complex task requests submitted by users, cloud computing task scheduling strategy should take into account economic benefits, resource utilization, enriching user experience and other factors in task scheduling and resource allocation. The basic flow of cloud computing task scheduling is shown in Figure 1.

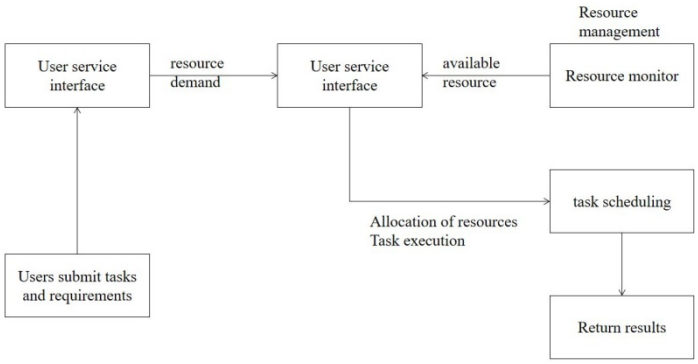

Figure 1 overview of cloud computing task scheduling

The basic principle of cloud computing is to distribute huge and complex storage and computing processing programs to a large number of distributed computers through the network, and provide corresponding application services, so as to help enterprises switch resources to needed applications, so as to enable users to access computers and storage systems according to their needs. In terms of the protection of information resources, "box computing" and "cloud computing" have their own ways and methods. "Box computing" is a kind of "dark protection" for the protection of information resources. This technology turns practice into theory for the protection of information resources, which is also information resources. However, "box computing" focuses on the introduction of theory, or even the protection of access rights. It protects customers' information resources in a way of "showing you but you can't see", The protection of information resources by "cloud computing" technology belongs to "explicit protection", because each company has its own data port, and people who don't know can't access it at all. Even in public information resources, there will be inaccessible information, or even those protected information will not be presented at all [7]. This protection is through "don't show you" To protect customers' information resources. This protection method is relatively strong, direct and effective. The starting points of "box computing" and "cloud computing" are different, so they have different ways to protect information. The former requires browsing, and the latter is a perfect information protection system [5].

6. Conclusions

By analyzing the existing scheduling work, this paper uses the five scheduling elements to comprehensively describe the essence and connotation of the scheduling problem, and illustrates the specific manifestation of each element with examples. Then, it makes a comprehensive analysis of the classical scheduling algorithms in traditional parallel distributed computing. At the same time, by analyzing the scheduling research in the current cloud computing environment, Combined with the new problems and new challenges brought by cloud computing, the development trend of cloud scheduling is pointed out. The above work can not only help quickly understand the classical scheduling algorithms and conclusions in parallel distributed computing, but also play a guiding role in the research of scheduling problems in eddy current search algorithm, and further promote the research and application development of cloud computing.

Acknowledgements

Research on budget constrained eddy current search intelligent optimal scheduling algorithm for Cloud Computing: Scientific research project of Hunan Provincial Department of education in 2019. Project (No:19C0964)

References

[1] Yuan Hao, Li Changbing. Cloud computing resource scheduling based on social force group intelligent optimization algorithm [J]. Computer science, 2015 (04): 206-208

[2] Hao Liang, Cui Gang, Qu Mingcheng, et al. Energy consumption resource scheduling optimization algorithm for cloud computing under cost constraints [J]. High tech communication, 2014 (5): 458-464

[3] Han Shuai, Liu Shulin. Impact of budget constraints on search engine revenue [J]. 2021 (2017-1): 18-21

[4] Jia Renfu, men Chunhua, Wang Hong. Application of heuristic search algorithm in optimal operation of pumping station [J]. Journal of Yangzhou University (agriculture and Life Sciences Edition), 2000, 21 (04): 63-65

[5] Zou Shengliang, Yang Jianjun. Research on Optimization Algorithm for solving dynamic job shop scheduling problem based on ILOG platform [J]. Dual use technology and products, 2010, 000 (003): 42-44

[6] Wu guanhan, Jia Weimin, Zhao Jianwei, et al. Design of multi UAV aided communication system under hybrid game mode based on Multi-Agent Reinforcement Learning [J]. Journal of electronics and information, 2021, 43:1-11

[7] Bian Qian. An improved new cuckoo search algorithm applied to the optimization of industrial flow scheduling problems. [J]. digital communication world, 2020, 000 (002): 83-84.