

Resource Scheduling Method of Meteorological Satellite Ground System Based on Load Balancing

Zhanyun Zhang^a, Manyun Lin, Xiangang Zhao, Lizi Xie, Lan Wei, Cunqun Fan^b

National Satellite Meteorological Centre, Beijing, China

^aemail:zhangxi@cma.gov.cn, ^bemail:fancq@cma.gov.cn

Keywords: meteorological satellite; ground application system; resource scheduling; load balancing

Abstract: Meteorological satellite data is large, and the corresponding data tasks are complex and diverse. Satellite ground application system makes the entire ground system resources applications some bottlenecks due to carrying more tasks. That how to improve the resource utilization of the whole system has become a key problem in how to allocate the resources of the satellite ground application system rationally. In this paper, a resource scheduling method for meteorological satellite terrestrial application system based on load balancing is proposed. Firstly, the resources are defined quantitatively. Then, the resource requirements are formally described by matrix. Finally, the reasonable scheduling of resources is realized by load balancing. The proposed method can realize the load balance of the system resources under the premise of guaranteeing the efficiency of task execution.

1. Introduction

Meteorological satellite ground application systems receive data from meteorological satellites and generate and distribute meteorological satellite data products of different levels and types according to different business needs. The amount of meteorological satellite data and a large number of products, satellite ground application system will face more tasks carrying, more computing, higher storage needs and a wider network resources. Based on the resource scheduling of the cloud satellite ground system can be achieved on the ground system of virtual resource scheduling, improve resource utilization. However, clouded satellite ground system resources are limited and can not be allocated to mission resources indefinitely. At present, cloud computing is increasingly concerned by the academic field and business areas, along with the explosive development of cloud computing, seeking cheap and computationally powerful platform is imminent. The scheduling strategy of virtual resources under cloud environment has become the core of cloud computing research, not only has very important theoretical value, but also has extremely important practical significance.

At present, the research on resource scheduling has made a lot of research results. Paper ^[1] introduced cloud architectures for biomedical informaticists who may wish to build applications using a cloud, and for investigators who want to share data with collaborators. Paper ^[2] considered the objective of minimizing the congestion and show that solving a subproblem about mapping a single workload to the physical graph essentially suffices to solve the general problem. Paper ^[3] proposed a novel VM images acquisition scheme based on multi-point collaboration (MPCA), which minimizes latency in fetching VM images. A method for energy efficient resource allocation was proposed in ^[4]. Paper ^[5] presented a general resource management architecture that uses the stable matching

framework to decouple policies from mechanisms when mapping virtual machines to physical servers.

2. System Resource Scheduling Method Based on Load Balancing

2.1. Resource Model Establishment

In recent years, many scholars have studied the programming model in cloud environment, most of them adopt Map/Reduce model. Map/Reduce is not only a distributed computing model for large-scale data processing, but also an efficient task scheduling model. The whole calculation process is done by master worker and user. Where the worker thread is equivalent to the central scheduler of the entire centralized system, completing the task of data segmentation task scheduling and equalization control; the worker thread receives tasks from the master thread, performs specific tasks such as data processing and operations, and The data is transferred to the corresponding storage location; user thread is the user thread.

In this paper, the virtual machine resource scheduling is divided into two levels: the primary scheduling is the cloud of user tasks and virtual machine resources matching problem, the user task and virtual machine resources to a certain optimization objectives for the principle of mapping, so that the user task in the appropriate; The secondary scheduling is the virtual machine and the host mapping process, so that the virtual machine in the appropriate host to create or migrate. This paper focuses on the first-level scheduling of virtual machines, that is, virtual machine resources and user tasks mapping problem. The optimal mapping between the task and the virtual machine is established with the optimal time span and load balancing. Define T_c for the time required to complete a task on a particular resource, then T_c can be expressed as follows.

$$T_c = t_w + t_s + t_e \quad (1)$$

Where t_w represents the task latency, t_s represents the task data transmission time, and t_e represents the execution time of the task on the resource.

2.2. Load Balanced Resource Scheduling.

Driven by demand, the number of virtual machines in the cloud is increasing, and the allocation of virtual machine resources has become a new challenge. A simple allocation strategy causes a virtual machine to bind a physical server for life, and a single attribute-based scheduling policy can cause uneven load, for example, a virtual machine that requires memory resources may be deployed on a physical server where the CPU resources are free, Scheduling efficiency and load balancing can lead to waste of resources or lack of resources. In the process of resource scheduling, this paper uses the virtual machine CPU load, memory load and network load index to dynamically respond to the current load of virtual machine resources to prevent load imbalance. In this paper, the resource requirements of the task, virtual machine resources to model, including processing needs, memory requirements and network bandwidth requirements. Each task can be represented by $S_t(N_t, M_t, C_t)$ to the resource requirements, where C is the processing resource requirement, M is the memory resource requirement, and N is the bandwidth resource requirement. Similarly, each virtual machine can also represent resource requirements by $V_v(N_v, M_v, C_v)$, where C represents processing resource requirements, M represents memory resource requirements, and N represents bandwidth resource requirements.

When allocating resources, they can be formally described by the matrix, the matrix unit values are set to 0 and 1, 1 for efficient allocation, and 0 for invalid allocation. The allocation matrix is as follows:

$$SV = \begin{bmatrix} ab_{11} & ab_{12} & \cdots & ab_{1m} \\ ab_{21} & ab_{22} & \cdots & ab_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ ab_{n1} & ab_{n2} & \cdots & ab_{nm} \end{bmatrix} \quad (2)$$

Where $ab_{ij} = 1$, then task i is considered to be effectively allocated to resource j for execution. $ab_{ij} = 0$ indicates that task i is not assigned to resource j .

According to the above analysis, the resource occupancy rate of virtual machine j can be expressed as follows:

$$V_j = \left(\frac{k_1 \sum_{i=1}^m (ab_{ji} N_i)}{V_{nj}} + \frac{k_2 \sum_{i=1}^m (ab_{ji} M_i)}{V_{mj}} + \frac{k_3 \sum_{i=1}^m (ab_{ji} C_i)}{V_{cj}} \right) \times 100\% \quad (3)$$

Among them, k_1 , k_2 , k_3 , respectively, said the network, storage and CPU weight coefficient, the weight coefficient can be set according to different system requirements. Under normal circumstances, $k_3 \geq k_1 + k_2$ and $k_1 + k_2 + k_3 = 1$. That is, the weight of the CPU will be set higher, CPU weight is generally greater than the sum of the network and the weight of storage.

The average resource occupancy rate for a virtual machine can be expressed as follows:

$$V_a = \frac{\sum_{j=1}^n V_{oj}}{n} \times 100\% \quad (4)$$

The standard deviation function of the virtual machine node resource allocation number is:

$$\delta(j) = \sqrt{\frac{\sum_{j=1}^n (V_{oj} - V_a)^2}{n}} \quad (5)$$

In the process of resource scheduling, we first detect the performance indicators of virtual machine resources, calculate the current resource occupancy rate and resource allocation standard deviation function, so as to understand the virtual machine load situation in real time, find more suitable virtual machine resources to reduce Virtual machine resource load, so that the resource request to deal with to achieve load balancing.

3. Simulation and Analysis

3.1. Simulation Establishment

In order to verify the effectiveness of the proposed algorithm, we implemented our algorithm on the Visual Studio 6.0 platform and compared it with the resource-based mapping method based on task weight and the resource mapping method based on task queuing time.

Simulation environment to establish the PC configuration: CPU Core i3-2310M 2.10GHz, RAM = 4GB, Windows 10 operating system. The experimental simulation system is written in C++ language. The mathematical functions involved in the experiment are calculated by calling the math.h math function library. The method proposed in this paper has no specific requirements for the tasks in the

cloud satellite application system. In other words, the satellite ground application system needs to deal with the task can be applied to the proposed method. In the experiment, we set up 100 target tasks, each task randomly allocates three resource requirements components and processing time, the resource demand component is in the range of [900, 1600] MIPS computing processing resources, [100, 500] G Storage resources, [8, 50] Mbps bandwidth resources, processing time [90, 600] s. Resource pool total resources for the 30000MIPS computing processing resources, 10T storage resources, 2000M bandwidth resources.

3.2. Analysis of Results

In the system, the task of multi-task requires a higher timeliness, that is, for a certain amount of task processing time to be as short as possible. Figure 1 shows the time required for the proposed algorithm (BLB) to process 100 tasks and the comparison with the task-based resource scheduling algorithm (BWA) and the queuing time-based resource scheduling algorithm (BQA).

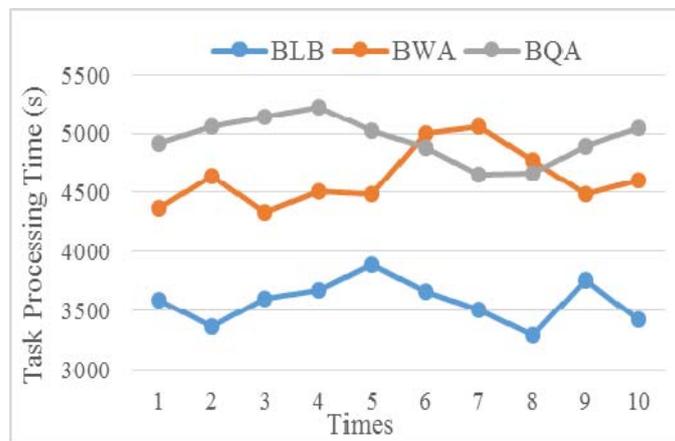


Fig. 1 Task processing time

Figure 2 shows the comparison of the resource utilization ratios of the 100 tasks with the proposed algorithm and the resource utilization of the resource scheduling method based on the task weight and the resource scheduling method based on the queuing time.

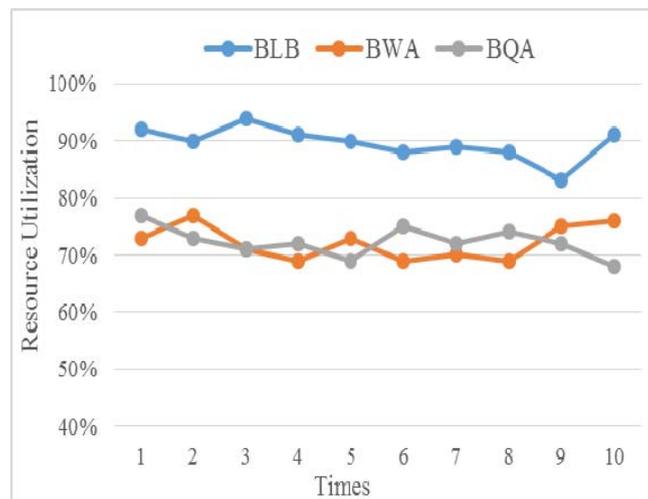


Fig. 2 Resource utilization

The experimental results can be seen that the proposed algorithm is superior to the contrast method in task processing time and resource utilization.

4. Conclusion

For the satellite ground application system resources for reasonable scheduling, thereby improving the overall system resource utilization. In this paper, a resource scheduling method for meteorological satellite terrestrial application system based on load balancing is proposed. Firstly, the resources are defined quantitatively. Then, the resource requirements are formally described by matrix. Finally, the reasonable scheduling of resources is realized by load balancing. The proposed method can realize the load balance of the system resources under the premise of guaranteeing the efficiency of task execution.

Acknowledgements

The work presented in this study is supported by National High-tech R&D Program (2011AA12A104).

References

- [1] A. Rosenthal, P. Mork, M. Li, J. Stanford, D. Koester, P. Reynolds, Cloud Computing: A New Business Paradigm for Biomedical Information Sharing, *Journal of Biomedical Informatics*, 2010, vol. 42, issue. 2, pp. 342-353.
- [2] N. Bansal, K. W. Lee, V. Nagarajan, M. Zafer, Minimum congestion mapping in a cloud, *Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed computing*, New York, USA, 2011, pp.267-276.
- [3] B. Huang, R. Lin, K. Peng, H. Zou, F. Yang, Minimizing Latency in Fetching Virtual Machine Images Based on Multi-Point Collaborative Approach, *Green Computing and Communications (GreenCom 2013)*, Beijing, China, Aug. 20-23, pp. 262-267.
- [4] A. Leivadeas, C. Papagianni, S. Papavassiliou, Energy Aware Networked Cloud Mapping, *2013 12th IEEE International Symposium on Network Computing and Applications (NCA)*, 2013, pp.195-202.
- [5] H. Xu, B. Li, A Versatile and Efficient Framework for Resource Management in the Cloud, *IEEE Transactions on Parallel & Distributed Systems*, vol.24, issue. 6, 2013, pp.1066-1076.