2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Reducing the Power Consumption of Servers with Bandwidth Consideration

Yu-Chang Lin
Dept of Electrical Engineering
TamKang University
Taipei, Taiwan
bearlaker34@gmail.com

Wei-Tsong Lee
Dept of Electrical Engineering
TamKang University
Taipei, Taiwan
wtlee@mail.tku.edu.tw

Jing-Yue Qiu
Dept of Electrical Engineering
TamKang University
Taipei, Taiwan
freedom4161@gmail.com

***Abstract*— In recent years, cloud computing systems become more and more mature and cloud computing system applications are becoming more widespread. Microsoft, Google, IBM, Amazon has developed applications for the cloud computing environment. The cloud computing environment like a large pool of resources, MapReduce distribute resources in this resource pool to achieve cloud computing. Hadoop MapReduce is a distributing cloud computing system, more and many user this cloud computing system. Green Mater[1] is a resource manager in a cloud computing system which is based on Hadoop MapReduce, Green Master[1] distribute jobs out to the virtual machines with higher computing performance rather than that with lower computing performance. Green Master[2] select virtual machines which have higher performance and assign tasks and reducing virtual machines achieves power-saving effect. This paper focuses on adding network bandwidth in Green Master, we consider the transmission speed of network bandwidth into Green Master and filter out three appropriate virtual machines, and those virtual machines have better performance than Green Master select.**

Keywords—Hadoop; MapReduce; Benchmark; Cloud Netwrok; Bandwidth

I. INTRODUCTION

MapReduce[2] is a programming model, more and more people use the cloud computing system. Hadoop MapReduce which is the evolution of the system from the Google cloud is an open source project. In the cloud computing system distribute files and assign tasks to a virtual machines or physical machines then acquire resources and services. These MapReduce functions are connected by a network to acquire computing resources which Master assigns. Therefor software algorithms, hardware performance and network transmission rate will impact on the operational efficiency of the distributed computing system. There are some solutions proposed in many papers, the effective management of servers in accordance with the schedule of work tasks choose Select server computing in the form of data and so on. For acquire maximum resources, usually turn on multiple machines to join into operation, that can make computing faster, but also produce relatively energy problem, because each server performance are different, so the operation efficiency is different. Each of virtual machines have different performance and task completion times are not same, the higher performance complete map function quickly, but in MapRreduce programming model must wait all machines finish, after all map function are finished then do Reduce function, that will cost a lot of time and energy to waiting lower performance virtual machine. Green Mater reducing the power consumption of the cloud system which is based on Hadoop MapReduce, Green Master will select three virtual machines which have higher performance and avoid using low-performance virtual machine. Green Master reduce use low-performance virtual machines but it does not have seriously affect in the overall system performance, and achieving the energy saving effect. This paper we consider the network data transmission rate to select the three appropriate virtual machines which have good performance and network bandwidth. In MapRduce process, when a job is distributed and assigned to do the operation on the specified virtual machine, the virtual machines scattered around the subject over the network to send data over these scattered, so the speed of the network bandwidth also affect the overall operation is complete.

II. RELATED WORKS

A. Hadoop

Hadoop MapReduce is a system which comes from Google MapReduce, which is a distributing system. Hadoop split file into several small files and distribute these small files to machines for parallel computation, then collecting result after processing is completed. There are three parts in Hadoop MapReduce, each are Master, Map and Reduce.

B. Master of Hadoop

Master like a controller in Hadoop, Master is an important node that cannot be replaced by other nodes. Master node improves the efficacy of cloud computing system. Master play manager role in Hadoop, Master has two important jobs, one is assigning the virtual machines to Map and Reduce, the other job is assigning tasks to each nodes. Master receives the command from user, and assigning tasks to task trackers. Task trackers will send a message to Master node, let Master known Work node status, and the status store of task tracker in database. The status of work node has three different types: Idel, In-processing and competed, according to the status Mater which VM can assign tasks, as show as in Fig 1 [1]

978-1-4799-5390-5/14 $31.00 © 2014 IEEE
DOI 10.1109/IIH-MSP.2014.168

IEEE computer society

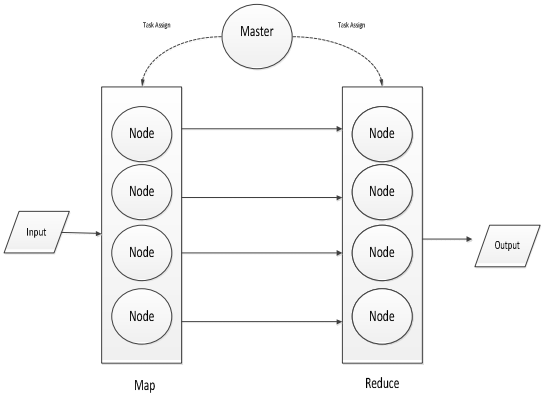

Fig. 1. MapReduce Architecture

C. MapReduce

Hadoop MapReduce is developed from Google MapReduce which is parallel computing. Hadoop MapReduce will separate big files into small files, and dispersed to do calculations on virtual machines. After Map and Reduce complete tasks, the result will be entire collection. There are three parts in MapReduce, each are Master, Map and Reduce. Map's work is computing those tasks into special format which are allocated to Map, t hose special format tasks that we call Intermediate Data. Using the Key/Value to do classification when Map compute word count. When Map's works are done, Reduce will get the message form Master, massage will tell the address of Intermediate Data to Reduce, then Reduce collect these data and combines the result according to Key/Value.

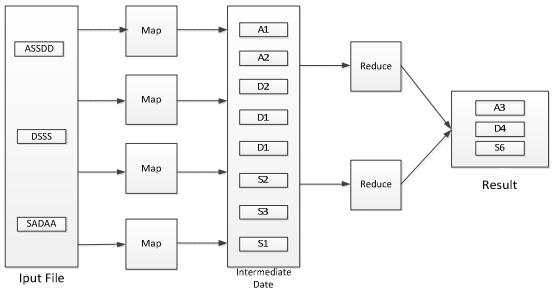

Fig. 2. Key/Value

D. Virtual Machine

It is a trend that applying VM technical on cloud system to provide resource share ability. Operating System-Level Virtualization allow a machine cut out multiple virtual machines , each of machines can install a operate system on them. Server virtualization is divided into two ways:

- Hosted Virtualization:

A virtualization software is installed on a host operating system to support machine virtualization is called Hosted Virtualization. For example, a computer can be installed a Windows operating system and VMware on it to support many different operating systems running on the same machine.

- Bare-Metal Virtualization

There is a single computer which has not been installed any operating system, but a virtualization software. In other words, the virtualization software is installed directly on hardware. For example, users can install VMware ESXi(VMM) directly on the hardware of a computer, we can use VMware vSphere client or the command to control the VMware ESXi, This way is called Bare-Metal Virtualization.

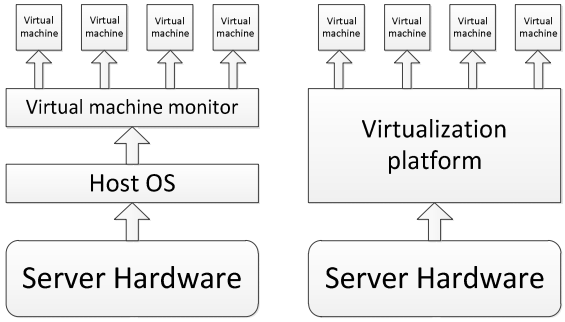

Fig. 3. Virtual Machine

E. Green Master

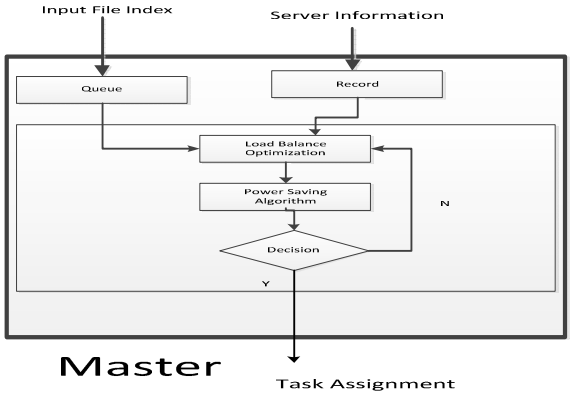

Fig. 4. Green Master Architecture

According to the results in [2], Green Master is a controller of a system which has efficient allocation process to save energy in limited resources. Different from Hadoop Master, Green Master System (GMS) applies a novel mechanism to assign tasks to virtual machines and thus reduce the power consumption of the cloud system. GMS choose three high-performance virtual

machines and assign the task according to its priority. In this way the system can avoid performance degradation and energy waste.

F. Benchmark Score

Benchmark[3] performance is mainly used to test the object, the performance rating of the machine provides users with a reference value relatively good or bad. VM Benchmark is new testing performance methods for virtual machines, VM Benchmark score can test virtual machine CPU, hard drive, memory. In Green Master use Benchmark score on several VM, and according to Benchmark scores sort of distinction between the performance of VM, which can provide cloud computing better performance, then Master will reference the VM's score to assign tasks.

III. IMPLEMENTATION

In this paper we reduce the power consumption of the Green Master System, by considering network bandwidth in job allocation. But GMS does not consider bandwidth into these three VMs, so let bandwidth inclusion of network bandwidth choice will produce new results.

A. Server Information

VM's information of benchmark score and available network of VM will be stored in server as show in figue 5.

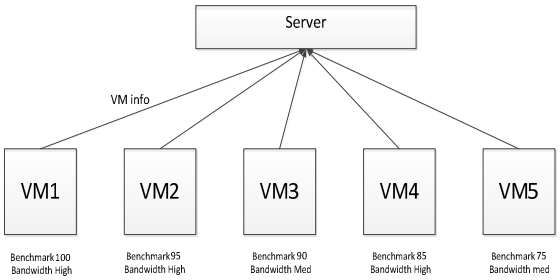

Fig. 5. Server Information

According to the results in [1], using three VMs to process the computation results in minimal impact on performance with maximum power saving. Therefor those VMs will be VM1, VM2 and VM3. In this paper, we take network bandwidth and VM benchmark score into consider. First we compare VM3 and VM4，if the VM3's benchmark score minus VM4's benchmark score is less then 15, and VM4's bandwidth is better than VM3, then VM4 will replace VM3 to join the work. VM3's benchmark score minus VM5's benchmark score is bigger than 15, although VM5 has better network bandwidth than VM3's. But as show in Figure 11, using VM5 to replace VM3, VM5's system consumption is worse than using VM3.

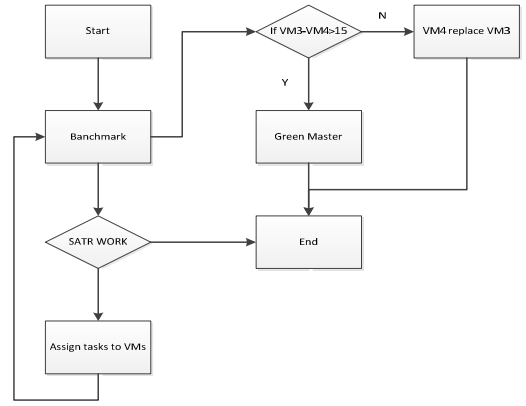

Fig. 6. Decide VM Replacement Process

IV. SIMULATION RESULT

In this paper, we have seven additional virtual machines. As show in Figure 5, there are seven different virtual machines are configured system resources. Master gets the network status of these seven virtual machine, the virtual machine will be credited each benchmark score and sort, and then we compare VM's benchmark and the network .

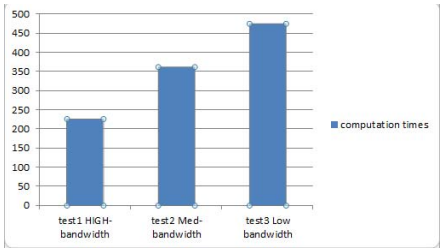

Fig. 7. Processing Time with Different Bandwidth

In figure 7, we use [1] test processing time with different bandwidth, we can compare the high-bandwidth and the low-bandwidth processing time, GMS with low-band need double time to finish the task process.

A. The processing time GMS under different available network Benchmark.

In figure 8, we use different files to do Green Master MapReduce, we can see that network bandwidth is a important factor, when input files is 2 GB, using the high-bandwidth's system time is faster than using low-bandwidth.

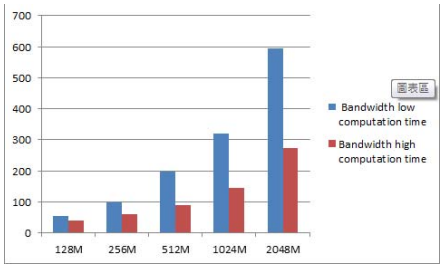

Fig. 8. GM Computation Time with Different Input

B. *Compare VM3 and VM4 the system consumptions and computation time*

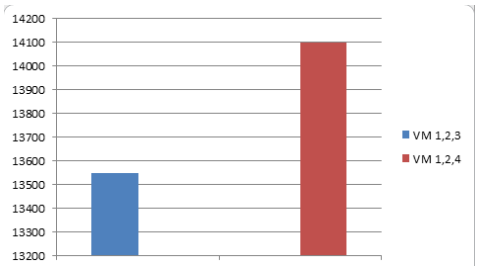

Fig. 9. VM3 and VM4 consumptions

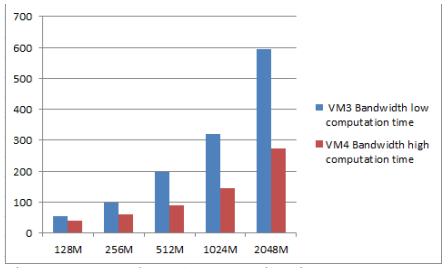

Fig 10. VM3 and VM4 computation time

As show in figure 9, VM4 system consumption cost more 550(w) than VM3, but VM'4 computation time is faster than VM3.

C. *Compare the total system consumption VM3 low- bandwidth and VM4 high-bandwidth.*

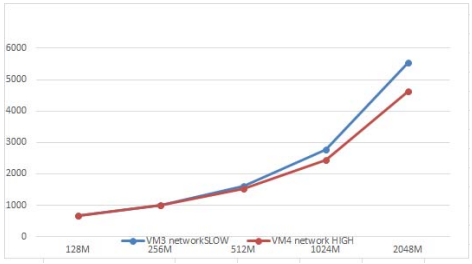

Fig. 11. System Consumption VM3 and VM4

We take VM3 and VM4 system consumptions in figure9 times their computation time and the result become VM3 and VM4 System Consumption(KJ). We can find out that using VM1, 2, 4 cost more consumption, but the total system consumption is better than VM1, 2, 3.

D. *Use VM5 to Replace VM3 system consumption*

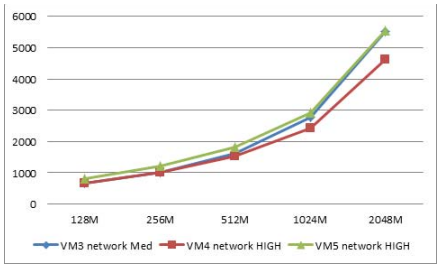

Fig. 12. Compare VM3 VM4 VM5 System Consumption.

In Figure 12, if we use VM5 to replace VM3, VM5 total system consumption are worse than VM3, because VM5's benchmark 75 that why its consumption cost more 3350(w) than VM3.

V. CONCLUSION

As show in our experimental results, VM's performance and network bandwidth both are important in cloud system. When Master start assign tasks to VM, choosing VM with high benchmark score and high-bandwidth will be the best options.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science and Technology, Taiwan, R.O.C., under the grant No. NSC 1102-2221-E-032-029-.

REFERENCES

[1] Wu, Ming-Zhi; Lin, Yu-Chang; Lee, Wei-Tsong; Lin, Yu-Sun; Liu, Fong-Hao "Green Master based on MapReduce Cluster", The 8th International Conference on Embedded and Multimedia Computing (EMC-13)

[2] *Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simpli_ed Data Processing on Large Clusters," OSDO 2004.*

[3] Xia Xie, Qingcha Chen, Wenzhi Cao, Pingpeng Yuan, Hai Jin, "Benchmark Object for Virtual Machines," 2010 Second International Workshop on Education Technology and Computer Science.

[4] *Lars Kolb, Andreas Thor, Erhard Rahm, "Load Balancing for MapReduce-based Entity Resolution," 2012 IEEE 28th International Conference on Data Engineering.*

[5] *Tao Zhu1,2, Chengchun Shu1, Haiyan Yu1, "Green Scheduling: A Scheduling Policy for Improving the Energy Efficiency of Fair Scheduler," 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies.*