

Storage Optimization for Energy-Saving Based on Hypergraph in Cloud Data Center

Xudong Chen¹ and Baomin Xu²

¹*School of Software Engineering, Beijing Jiaotong University,
Beijing 100044, China
chenxd@bjtu.edu.cn*

²*School of Computer and Information Technology, Beijing Jiaotong University
Beijing 100044, China
2889130209@qq.com*

Abstract

With the rapid growth of data-intensive computing, cloud data center using MapReduce framework are increasingly popular. The energy consumption has become the most important problem in cloud data center which is based on MapReduce framework. MapReduce framework is typically deployed in large-scale clusters with thousands of nodes. The replicas are used to provide data redundancy. In this paper, we proposed a hypergraph-based power saving algorithm of data replica, which is based on the analysis of default placement strategy in Hadoop. The experimental results show that the proposed algorithm has certain advantages such as fault tolerance and availability of data blocks. Also, the energy efficiency is better than CS-k algorithm in the same operating environment.

Keywords: *Data center; energy efficiency; hypergraph; replica.*

1. Introduction

In recent years, new discoveries and new business fields are established on the basis of big data storage and analysis. The rise of Internet enterprise scale needs higher request to the server hosting and low cost of professional services. While the investment and management cost of computer equipment prohibitive to many small and medium enterprises. Traditional data center and distributed architecture cannot adapt to the footsteps of modern enterprise development. The industry needs a new architecture and commercial pattern to solve this bottleneck.

The architecture and service model of traditional IDC (Internet Data Center) cannot adapt to the requirement of modern enterprises' effectiveness. New data centers established on cloud computing technology [9, 10] have high scalability, high reliability, easy to share and very low technical threshold which a lack of past data center. The new data centers reduce the burden of customers and make them focus on their core business development.

Due to the strong growth of computing power demand, all kinds of small and large data center are constructed rapidly. Studies show that data center in America consumes 61 billion kwh in 2006 which accounted for 1.5% of the total electricity of America. The energy can meet 5.8 million families of America and cost about US \$4500000000[1]. According to the report of the EPA (Environmental Protection Agency), we can reduce at least 150 to 4.7 million tons of carbon dioxide by reducing energy consumption [2]. Despite the technology progress, there still are 6% data center servers which have low utilization rate. Advocate green computing, build green data center has become an extremely urgent and important research direction.

2. Related Work

The energy consumption in data center mainly come from the computing processing, storage, network transmission, cooling system [3], and power protection etc. At present, the research focuses on hardware, software architecture, and the idle node processing. The hardware mainly includes equipment update and refrigeration. Software architecture focuses on the data exchange and memory. And the third point is about how to improve the utilization rate of nodes and reduce energy consumption.

In the aspect of hardware equipment, data center contains thousands of servers typically. It will need huge amounts of energy when a large number of servers running at the same time. We could replace hardware devices and purchase more cost-effective low power CPU and energy-saving hard disk for reducing power consumption. In the aspect of computer room designing, due to the data center server typically placed in a fixed room and intensive distribution, it will produce a lot of heat. People need to take a variety of cooling measures to keep it in a suitable temperature and humidity. The device itself also becomes an important source of heat and energy consumption.

In the aspect of software architecture, Dzmitry etc.[4] did a lot of work on the optimization of data center network and data exchange. The authors analyse data flow and the main source of energy consumption and put forward energy efficiency and network awareness scheduling method using the three-layer architecture of data center. This method builds a weighted metrics on the server level, frame level and module level. It chooses the best adapted computing power and avoids the formation of hot spot within the data center on the basis of load balancing and considering communication cost.

Many data centers do not need keep all compute nodes working all the time. When user is less especially in the night and the morning, the load of many data center is quite low. And the idle CPU power consumption is more than half of the work state, so a lot of computing resources and energy is wasted. Under the premise of ensuring data integrity and user QoS (Quality of Service, Quality of Service), we need to improve the existing server utilization and reduce energy loss as far as possible. For single node and the chip, we can use DVFS technology change the CPU frequency and voltage. For the data center which order of magnitude is in thousands of millions, there are two main methods to improve server utilization for now. One called CS (Covering Set, namely the cover Set), it will choose some nodes among all servers and make full use of these node while shut down the remaining nodes. Another called AIS (All in the Strategy, namely overall Strategy), the task will be assigned to all nodes and shut down the server as soon as possible after the completion of all jobs. To achieve better energy-saving effect, the above two methods are usually combined with the DVFS technology and virtual technology.

Jacob Leverich[5] proposed the definition of CS (Covering Subset, namely covering subsets) firstly. The author modified the existing data replicas algorithm of Hadoop so that the CS node contains only one replica of data items. The method is more suitable for the clusters which have more stable computing needs. Harnik D[6] used the combination of random and greedy algorithm to select the cover set.

Jinoh Kim[7] promoted the CS method and proposed a CS-k algorithm to find multi-level nodes. This algorithm can find cover and k copies of the data set under the heterogeneous environment. It also studies the problems of heterogeneous computing cluster and balancing energy consumption.

3. Analysis of Replicas Storage Problem

3.1 Storage Mode of the Replicas

For data center and cluster, data storage and compute are equally important. Server and physical hard disk are the data storage medium. The damage probability of them is not big. But once failure happened, the loss of data and suspended services will cause a great

economic loss. The good data backup mechanism is a guarantee to ensure data availability. At present, the RAID and redundant data copies are used in practice.

Distributed file system such as GFS, HDFS mainly use data replicas to provide data redundancy. The technology can improve performance and data availability. These system can develop different levels of data redundancy scheme according to user requirements. However, data replicas mean more storage space.

The MapReduce framework adopts multiple copies backup strategy based on frame perception. It brings great pressure to the data storage of data center. In order to reduce data copies to reduce storage costs, Facebook presented HDFS-RAID technology on the basis of HDFS. In decreasing the number of copies of data at the same time, the use of RAID error correcting code and check code in solving the problem of data availability.

3.2 Optimization of Replicas

The data block in the HDFS has 3 default copies. Number 1 is stored in the compute nodes generally. Number 2 is stored in the server which is different to number 1 and at the different rack. Number3 is stored in the server which is different to number1 but at the same rack. We have a third replicas of the data stored in the same server and two thirds of the replicas of the data stored in the same rack. And remaining copies were distributed randomly in the other servers. Set data blocks A and B belong to Rack1 and Rack2. Two copies of their distribution are shown in Figure 1.

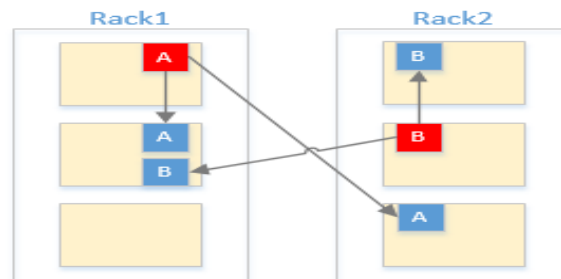


Figure 1. The Placement Policy of Data replica in Hadoop

The default storage strategy of Hadoop is considering the reliability (data blocks are stored in two different frame), write bandwidth (frame faster data transmission between internal server, writes only converter through a network), reading performance (can choose to read data from the two racks), distribution (users only need to write data block into the local frame at a time). If replica factor is greater than 3, the redundant copies are stored in randomly selected servers in the cluster. Master/Namenode adopt the method of average distribution when allocating data blocks. This architecture allows the nodes in an open position in the cluster before finish all tasks. This greatly increased energy consumption. To achieve the goal of energy saving, we need to take certain measures to optimize replicas storage and process data within the cluster so as to close the free server.

4. Our Algorithm

Hypergraph has great advantages in multidimensional data representation and processing. For example it can easily express someone have multiple identities, or many people participate in an event. It is more accessible in terms of logic. In this paper, hypergraph model can easily express server nodes and the relationship among data replicas.

4.1 Hypergraph Construction

Based on the definition of hypergraph and Hadoop storage strategy, the hypergraph H is defined as follows:

$$H = \langle V, E \rangle$$

where $V = \{v_1, v_2, \dots, v_n\}$ is the set of physical servers, $E = \{e_1, e_2, \dots, e_m\}$ is the data set of replicas.

4.2 Algorithm Description

The traditional greedy algorithm is not suitable for our research. In this paper, we add node resource load limit in greedy algorithm. The novel algorithm SOPR is divided into reduction of copies, server status labelling and dynamic migration.

(1) Copies Reduction

- 1) Computing the degree of each vertex of hypergraph H ;
- 2) According to the descending of the sum of vertex degree, the descending of the sum of the remaining MIPS (Million Instructions Per Second) and the power consumption, each rack is added to the rack queue;
- 3) According to the greedy algorithm, take out the front racks in the sequence of nodes in turn and delete nodes connected over the edge of the ear (labelled delete logo).

(2) Server Status Label

According to server status label, the reserved servers are divided into four kinds. Namely, set the status of the server be 2, which the free load more than 50%; Set the status of the server be 1, which the free load less than 50% and CPU load greater than 70%; Set the status of the rest server within the queue be 0; Set the status of the rest servers outside the queue be 3.

(3) Dynamic Replicas Migration

Sort the server queue according to the value of the state, remaining CPU's Mips, free storage of the server. The computational tasks will be migrated to the pre-reserved server from other server. Return the Number of replicas to state which in accordance with three copies of the MapReduce framework after the migration.

5. Results and Analysis

Jinoh Kim [7] proposed a Multi-level Node Set Discovery algorithm CS – k (k is the number of nodes to cover), a set of nodes that contain at least k replicas of the data blocks. The algorithm uses greedy heuristic method and calculates a new set during iteration. The selected set must contains k-elements that have not yet been covered as much as possible. CS-1 algorithm and CS-2 algorithm are the comparisons to our algorithm, which contain at least 1 or 2 replicas of the data blocks.

Let $S = (V_1, V_2, \dots, V_m)$, the steps of CS-k algorithm are as follows:

- 1) Select a vertex V_1 that covers the most data blocks in S .
- 2) Select a vertex V_2 that covers the most data blocks in $S_1 = S - S(V_1)$.
- 3) Select a vertex V_3 whose degree is the biggest in $S_2 = S_1 - S_1(v_2)$
- 4) Loop steps 1-3 until cover set $S' = (V_1, V_2, \dots, V_n)$ covers all the data blocks at least k.
- 5) Close the nodes that are not in the cover set.

In our experiment, there are 20 servers in resource pool. They belong to 5 racks. We simulate the server cluster that has 5 racks and every rack has 4 Host. There are two kinds of Host. The configuration parameters of CPU, memory, hard disk, bandwidth and power consumption are shown in Table 1. The Host0 is in simple Rack0 and Rack1, the Host1 is

in Rack2, Rack2 and Rack4. There are 20 tasks in experiment and each of them is assigned to the virtual machines of the 20 servers.

Because at the end of running of SOPR algorithm, cluster returns to the state of three data copies, so the cover degree of SOPR algorithm is higher. The result is shown in Figure 2, SOPR algorithm can cover all the Cloudlet and Replica with the need of 3 racks and 12 Hosts.

Table 1. Parameters of Hosts

Host	CPU(mips)	RAM(mb)	Storage(gb)	Bm(mb)	$P^{(i)}(W)$	$P^{(p)}(W)$	$P^{(s)}(W)$
0	1000	1024	100	1000	150	250	10
1	500	512	40	1000	55	100	7

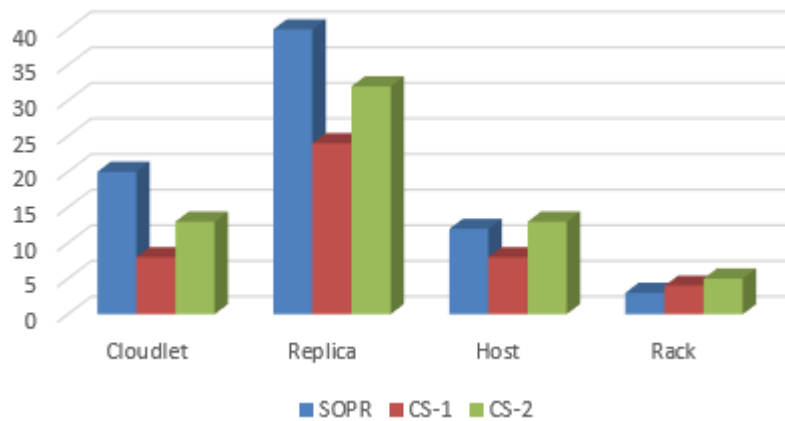


Figure 2. The Result of SOPR, CS-1 and CS-2

As is shown in Figure 2, Three algorithms have good performance in energy saving and can save 50%-80% energy. High performance hosts are selected in SOPR algorithm and CS-2 algorithm, so the energy consumption is much higher than CS-1. But considering the data availability, CS-1 algorithm has more advantages. Compared with CS-2, the coverage level of SOPR algorithm is more thorough. It uses less number of Hosts and its energy consumption is low.

The energy consumption of the algorithm of SOPR, CS-1 and CS-2 is showed in Figure3.

In addition, because of considering the running speed of CPU, the cover sets that SOPR algorithm generated are more likely to contain CPU servers that have higher computing performance, and it also has high power consumption. Therefore, SOPR algorithm can realize the purpose of energy saving of the cover set. SOPR algorithm follows SLA agreement basically so that it ensures the availability of the replicas of data. It is a good energy saving algorithm relatively.

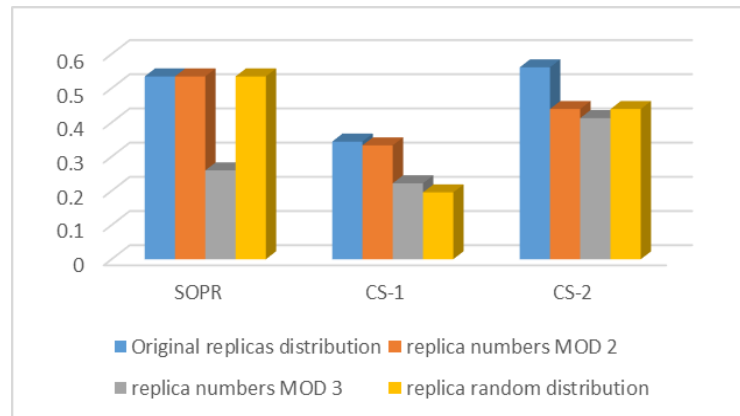


Figure 3. Power of Algorithms (W)

6. Conclusions

In this paper, our work is as follows: (1) converted the energy efficiency problem into a set covering problem and abstracts it as Hypergraph; (2) proposed a replicas dynamic migration algorithm considering CPU efficiency and deal with the transversal of hypergraph properly; (3) implemented our algorithm SOPR on the CloudSim. The experimental result shows that the improved algorithm has good stability and reaches a high level of energy efficiency while keeps the data availability.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC 61370060)

References

- [1] EPA Report on Server and Data Center Energy Efficiency, Public Law109~431, U.S. Environmental Protection Agency, ENERGY STAR Program <http://www.energystar.gov>.
- [2] J. M. Kaplan, W. Forrest and N. Kindler, "Revolutionizing data center energy efficiency", McKinsey & Company, Tech. Rep, (2008).
- [3] K. H. Kim, A. Beloglazov and R. Buyya, "Power-aware provisioning of cloud resources for real-time services", Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science. ACM, (2009).
- [4] D. Kliazovich, P. Bouvry and S. U. Khan, "DENS: data center energy-efficient network-aware scheduling", Green Computing and Communications (GreenCom. 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCoM), (2010).
- [5] J. Leverich and C. Kozyrakis, "On the energy (in) efficiency of hadoop clusters", ACM SIGOPS Operating Systems Review, vol.44, no.1, (2010), pp.61-65.
- [6] D. Harnik, D. Naor and Segall I, "Low power mode in cloud storage systems", Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on, (2009).
- [7] J. Kim, J. Chou and D. Rotem, "Energy Proportionality and Performance in Data Parallel Computing Clusters", Scientific and Statistical Database Management, (2011).
- [8] B. Xu, N. Wang and C. Li, "A cloud computing infrastructure on heterogeneous computing resources", Journal of Computers, vol.6, no.8, (2011), pp.1789-1796.
- [9] B. Xu, C. Zhao, E. Hu and B. Hu, "Job Scheduling Algorithm Based on Berger Model in Cloud environment", Advances in Engineering Software, vol.42, no.7, (2011), pp.419-425.

Authors



Chen Xudong, he is an associate professor in School of Software Engineering, Beijing Jiaotong University, China. His interests include network application, cloud computing technology and software engineering.



Xu Baomin, he is an associate professor in School of Computer and Information Technology, Beijing Jiaotong University, China. His research interests include Large Scale Data Analysis Based-on Cloud Computing, and Complex Network (Link Prediction, Community Detection). Until now he has published more than 50 research papers in referred journals and proceedings.

