

Research on Energy Efficiency Evaluation in the Cloud

KeHan^{1,a}, Xiaobo Cai^{2*,b}, XinZhang^{3,c}, Chaogang Wang^{3,d}

¹ Kunming Metallurgy College, Kunming 650300, China

² Yunnan Agricultural University, Kunming 650091, China

³ School of Humanities, Yunnan College of Business Management, Kunming 650106, China

^a hanke81@qq.com ^b xbcai918@qq.com, ^c 54332254@qq.com ^d 39266097@qq.com

Keywords: HPC; Cloud computing; Benchmark ;Energy efficiency

Abstract .High energy consumption issue is one of the major obstacles for the new cloud era. The energy efficient cloud environment with high performance and low power consumption has received extensive attention. In this paper, the definition of energy efficiency of cloud and the principles for implementing high energy efficiency cloud have been given. The mathematical expression and the measuring and calculation approaches are deduced. Besides, the key technology ,the problems and challenges are summarized and analyzed at last.

1 Introduction

Cloud computing has recently emerged as a new paradigm for hosting and delivering services over the Internet. With the advent of cloud computing era, the more computational resources and storage resources are concentrated in the clouds; the high energy consumption of the data center is becoming a prominent issue. In 2011, Professor Jonathan Koomey of Stanford University released a report that from 2005 to 2010 the data center consumption accounts for 2% of the total capacity of the US power grid. High energy consumption has become one of the most serious problems of cloud computing systems. So how to management energy consumption of Internet data centers(IDC) for users with quality of service (QoS) guarantees and higher performance, in other words, how to realizing the energy-efficient green cloud system is currently a major challenge for the cloud computing. In order to achieve the energy-efficient green cloud environment, we should find baseline assessment criteria to measure the energy efficiency of cloud systems first. Intuitively, a comprehensive evaluation of energy and efficiency can be a good measure of the relationship between performance and power consumption. Currently we often unify the both and choose specific energy, efficiency and operational metrics to evaluate the energy efficiency of cloud system on academic research and commercial applications.

2 The definition of energy efficiency

Currently studies for energy efficiency primarily according to the view of physics, refers to the ratio efficient energy consumption to the amount of energy actually consumed. Energy efficiency can be evaluated by two kinds of metrics by "energy" and "efficiency", expressed as a ratio of "Performance" to "power" per unit time of the system. Energy efficiency has been shown as an optimal way in cost-effective and energy consumption, while providing significant environmental benefits. In the cloud computing environment, from a cloud service provider's perspective, energy efficiency is an effective description of the cloud system equipment that energy consumption consumed. Understanding data center energy utilization equipment, energy-saving programs in favor of research for the cloud environment in the direction of the hardware and software design. From the consumer perspective, energy efficiency refers to the service for the end user to the total amount of consumed energy ratio, i.e. the ratio of the calculated energy. The so-called "energy efficiency" means to provide the same energy service with less energy input, or to provide more performance in the same proportion of energy consumption.

3 processing mode of Energy efficiency

To achieve the high energy-efficient cloud systems, we generally use two processing modes: First, discuss the energy management techniques such as sleep technology on the host computer and the level, DVFS, software optimization and hardware performance improvement, reducing the amount of resources used to reduce power consumption. Second, on the macro level, using a resource management system (RMS) in the data center, such as green resource management, thermal management, workload consolidation, task scheduling etc. In addition, the virtualization technology enables two modes of energy efficiency improvement more obvious, such as virtual machine technology, virtual machine migration as well as in a cloud environment. The traditional approach is to calculate the overall sum of various components of the energy efficiency and energy efficiency.

Energy efficiency can be described by the ratio of the system per unit time "performance" to "power" in the cloud environment. Performance can be calculated by the execution time or workload of running applications and computing tasks, while the energy consumption is described as energy consumed by units of time of the system. Most of the existing studies are used to measure electrical energy consumption Joule. Currently energy efficiency model and calculation method on cloud systems primarily in the following categories:

For the data center, green grid defined Power Usage Effectiveness (PUE) is the ratio of the energy used by the IT load to all data centers consumed as Data Center Infrastructure Efficiency (DCiE). Supercomputer divided the energy-efficient supercomputer calculation into two parts, performance parameters MFLOPS (millions of floating point calculations), power consumption unit measured by watts power (Watt) value, the final result is the ratio "MFLOPS / Watt", indicating the energy efficiency value. Literature [1] defined the completed computation by processor at 1GHz frequency in 1s. [2] proposed a definition of an atomic task unit, according to algorithm design task unit, such as sorting algorithms, this task can be defined to sort 100 amount of data as one unit, or define a transaction in the database system is a task unit or a user requests and responses as a task unit. Literature [3] proposed the ratio of QoS / energy consumption as a measure of the energy efficiency of cloud systems. Energy efficiency model has two parts, a lot of research proposed measurement model and performance measurement methods.

(1) energy measurement

The value of power consumption is usually adopted wattage units; most existing studies use the "Joule" to measure energy consumption. Power consumption per unit time is the size of J / S, i.e., watts (W). A typical rack server energy consumption [4] as shown in Table 1.

Component	Peak power	Count	Total	Percentage
CPU	40 W	2	80 W	37.6 %
Memory	9 W	4	36 W	16.9 %
Disk	12 W	1	12 W	5.6 %
PCI slots	25 W	2	50 W	23.5 %
Motherboard	25 W	1	25 W	11.7 %
Fan	10 W	1	10 W	4.7 %
System total			213 W	

Table I. Component peak power breakdown for a typical server.

(2) The performance test

Typically performance inspection is by measuring the number of running applications or computing tasks, or by the execution time or the workload. so the performance can describe * / S units. Literature [5] measured performance by the execution time of the tasks to service. [1] used the CPU usage and frequency to measure the size of a specific CPU load completed.

4 The key technology

The so-called "energy efficiency" means to provide the same service even more energy with less energy input.

4.1 Energy-saving technology

Saving mechanisms of power management in different ways can be divided into dynamic power management (dynamic power management, DPM) technology and static power management (static power management, SPM); according to the stage can be divided into closed / open technology (resource hibernation), dynamic voltage / frequency scaling (dynamic voltage and frequency scaling, DVFS) techniques and virtual machine (virtualization) technology categories, the former mainly lower idle energy consumption, the latter two mainly perform lower power consumption. The main premise of which is that the load condition DPM cloud systems are faced with constantly changing with time, it allows the current performance requirements for dynamic adjustment of power state, using DPM the main strategy has DVFS technology and virtualization. In contrast, SPM mainly take advantage of efficient hardware, such as CPU, hard disk storage, network equipment, UPS and energy to provide equipment, structural changes in the equipment can often reduce energy consumption.

4.2 Optimize the energy efficiency with QoS constraints

How to achieve energy-efficient cloud QoS guarantee, current research using constraints to meet the conditions for virtual machine deployment modeling, literature [6] constraints for users of the service level agreement, in order to achieve the purpose of maximizing energy savings. The basic idea is to maximize the number of physical machines idle, idle through close physical hosts to save energy. Literature [7] proposed a resource assessment model entropy optimization and dynamic weighting meet user QoS and system resources to maximize and achieve optimal scheduling, reduce energy consumption to achieve load balancing and improve system utilization based screening. The case of the literature [8] proposed an approach to meet customer requirements in the cloud layer SLA computing environment while minimizing energy consumption, and complete the deployment of cloud resources, making the overall system power consumption to a minimum. Literature [9] proposed the next best way to solve the energy problem is to increase energy efficiency or reduce energy consumption, the authors propose ACES-- automatic server configuration systems, data centers achieve load requirements decrease energy consumption, improve energy efficiency in order to achieve system purposes. Literature [10] Similarly to reduce data center energy consumption as a starting point, a new online algorithm, the algorithm can dynamically turn off the data center of the low load server dynamically to adjust the load distribution of the data center, in order to reduce the overall data center energy consumption.

4.3 task scheduling research based on energy efficiency

Cloud computing resources are provided to users by virtualization ways. The deployment of virtual machine resources under the cloud model is in fact a combination of a multi-objective optimization problem. To select multiple unrelated based additive measure QoS resource scheduling strategy is proved to be a NP problem which usually requiring the use of heuristic or intelligent optimization algorithms to solve. The traditional batch job scheduling to minimize the task execution time heuristic algorithms include ant colony algorithm, Max-min, Min-min, genetic algorithms. These algorithms usually optimize the task execution time, the target load balancing; scheduling system allows obtaining good performance and energy consumption. Energy-efficient scheduling strategies include: application load-aware, adaptive perceived economic costs, server clustering information including configuration, energy consumption and temperature status sensing, thermal management systems and temperature-aware load configuration, application independent research and establish a performance model, computer site load balancing.

Game theory can be an effective tool for network traffic congestion control, routing protocol, P2P incentive mechanism design and achieved good results. RanjanPal and others on a given situation of IaaS provider and QoS price, there exists a unique optimal Nash equilibrium [11]. AshrafA1 Daoud Stackelberg leader, who used the game for large data centers in order to maximize their effectiveness and conduct modeling demand pricing model based on the analysis of the data center to obtain optimal pricing strategies [12]. Jorge Londono, who deployed fixed physical resources with the introduction of the game theory in the cloud, but seeking the Nash equilibrium is

an NP-hard problem, the authors adopted a simplified, practical driving mix of game to prove its existence of Nash equilibrium [13].

5 problems and challenges for the Efficient cloud

First, the energy efficiency index have a wide range of influence in cloud computing. On the one hand, the application in the cloud computing architecture deployed in performance, latency, uptime normal service has strict requirements. How to ensure the quality of service requirements, while consider to optimize the energy efficiency of cloud computing system performance? on the other hand, the "cloud" service principle is user access to resources on demand. how to effectively configure cloud resources, with system performance guarantee, enables the system to maximum efficiency, and maximize the interests of service providers. Cloud computing model using a relatively centralized resource providing distributed services, different cloud service providers and users have different QoS requirements and cost requirements, different business types have different performance characteristics indicators. Faced with different providers, users and business types, how to get an effective assessment of energy-efficient cloud environment, how to quantify the cloud computing system and efficient "degree"?

Secondly, the cloud computing system has a large number of heterogeneous nodes, has complex environment variable, nonlinear and various uncertain interference factors. It is difficult to get an accurate mathematical model related to energy consumption, performance and QoS. The modeling errors and mutable environment usually resulting in energy efficiency value showing uncertainty in Existing studies. The relationship of power, performance and QoS is very complex, close coupling, interaction and mutual restraint. Reducing energy consumption is often accompanied by the loss of performance and quality of service degradation, how to clear the trade-off relationship between the three and the impact of energy efficiency-related factors?

Finally, virtualization is the cornerstone of cloud computing. Theoretically, highly virtualized cloud computing can generate less heat and reduce the energy cost of the system. In fact, server consolidation and virtualization often leads to a decline in throughput and other performance issues, but performance degradation will lead to the increase of the time required to complete the task, which will lead to cloud computing data centers' energy efficiency is uncertain. In addition, under the cloud computing model based on the quantitative relationship between energy efficiency and system performance analysis of QoS, how to design a rational mechanism to optimize resource deployment, ultimately meet different QoS requirements and SLA rating system with maximum energy efficiency.

Acknowledgement

This research was supported by with the nature and science fund from YunNan province ministry of education of China (No.: 2014Z076 No.:2014Y494) .

References

- [1] SONG Jie, LI Tian-Tian, YAN Zhen-Xing, NA Jun, ZHU Zhi-Liang Energy-Efficiency Model and Measuring Approach for Cloud Computing [J]. Journal of Software, 2012, 23(2): 200-214.(in Chinese)
- [2] SONG Jie,HOU Hong-ying, WANG Zhi, ZHU Zhi-liang. Improved energy-efficiency measurement model for cloud computing[J]. Journal of Zhe jiang University (Engineering Science) 2013, 1: 006. (in Chinese)
- [3]Bahsoon R. Green cloud:Towards a framework for dynamic self-optimization of power and dependability requirements in green cloud architectures. In:Babar MA, gORTONi,eds.proc.of the 4 th European Conf. on Software Architecture(ESCA 2010). Copenhagen,2010.510-514.
- [4] Fan X, Weber W D, Barroso L A. Power provisioning for a warehouse-sized computer[C]//ACM SIGARCH Computer Architecture News. ACM, 2007, 35(2): 13-23.

- [5] Kumar K, Lu Y H. Cloud computing for mobile users: Can offloading computation save energy?[J]. Computer, 2010, 43(4): 51-56.
- [6] Van H N, Tran F D, Menaud J M. Performance and power management for cloud infrastructures[C]//Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on. IEEE, 2010: 329-336.
- [7] Virtual Resource Evaluation Model Based on Entropy Optimized and Dynamic Weighted in Cloud Computing [J]. Journal of Software, 2013, 24(8).
- [8] Goudarzi H, Ghasemazar M, Pedram M. "SLA-based Optimization of Power and Migration Cost in Cloud Computing"[C]. 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2012: 172-179.
- [9] Guenter B, Jain N, Williams C. "Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning"[C]. 2011 Proceedings of IEEE INFOCOM,2011: 1332-1340.
- [10] Lin M, Wierman A, Andrew L L H, et al. "Dynamic right-sizing for power-proportional data centers"[C]. 2011 Proceedings of IEEE INFOCOM, 2011: 1098-1106.
- [11] Ranjan Pal, Pan Hui. "On the Economics of Cloud Markets"[J]. eprint arXiv: 1103.0045, Mar,2011.
- [12] Ashraf A1 Daoud, Sachin Agarwal, Tansu Alpcan. "Brief announcement: cloud computing games: pricing services of large data center"[C]. In the Proceedings of the 23rd international conference on Distributed computing, 2009.
- [13] Jorge Londono, Azer Bestavros, Shang-Hua Teng. "Colocation games: and their application to distributed resource management"[C]. In the Proceedings of the 2009 conference on Hot topics in cloud computing, 2009.