Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 1 (2017) pp. 121-128

http://www.ripublication.com

Energy Efficient Data Center: A systematic literature review

Navneet Singh[1] and Dr. Jatinder Singh Bal[2]

[1]Pursuing Ph.D.(Computer Application), Sant Baba Bhag Singh University, Jalandhar, Punjab, India.

[2] Advisors to Chancellor, Sant Baba Bhag Singh University, Jalandhar, Punjab, India.

Abstract

Objectives: We expect all electronic gadgetry to be 'energy efficient' to possibly achievable limits. So the slogan should be green cloud computing.

Methods: Data center known to consume lot of electric power. Here we shall discuss some of technique that are used to achieve this goal. In Server Power Management we have Independent voltage scaling [IVS], Coordinated voltage scaling [CVS], Vary-On Vary-Off [VOVO], Combined policy and Coordinated Policy. Then we have Virtualization, Live Migration and Energy Aware Profiling.

Findings: Broadly speaking cloud computing is nothing but a highly 'utilitarian' orientation of IT services where users benefited on a pay-as-you go basis. In a way it enables the hosting of pervasive applications from consumer, scientific and business domain. We expect all electronic gadgetry to be 'energy efficient' to possibly achievable limits. So our data centers hosting cloud application must be cost effective and the same time should avoid undue burden of carbon footprint. So the slogan should be green cloud computing. In this paper I shall discuss various techniques that help in making our data center energy efficient.

Improvements: The only drawback with cloud computing is that it is a notorious power guzzlers and call for a stringent 'energy efficiency' regime.

Keywords: Energy efficiency, Green cloud, Cloud computing, Resource management, Virtualization, Dynamic consolidation

1. INTRODUCTION

The concept of cc was first propounded by John McCathy way back in 1961, wherein he envisaged the idea of computer time sharing technology. His ideas and concept grew with time to become a reality. As they say nothing is beyond human contrivance and utilitarian aspects always egg man to strive for excellence. Today computing power and even specific applications could be sold through the utility business model like water and electricity and made available on-demand in metered way. So what cathey faintly dreamed about half a century ago took the shape of internet based computing and has been christened as cloud computing. Cloud Computing is a relatively new technology and aims to offer "utility based IT services".[1]

Like other public utility services wiz water supply, electricity, cooking gas and telephone etc. cloud computing is also considered a utility service which can be availed of on demand. It can also be rented out or charged in a metered way. Our modern society is increasingly relying on the collection, processing and sharing of digital information.[2]

Cloud computing is an arena that is ruling the world of information technology.[3] If we reflect and closely observe we find that the cluster computing of the last decade of the twentieth century followed by grid computing were earlier version of cloud computing. From 2002 onwards it started carving a niche of it's own in the mammoth edifice of information technology. Now it has established itself as a strong arm of society with added advantages of cost effectiveness pliability, easier metering and optimum utilization of media in a given time space. These distinguishing features of cloud computing are sure to attract industry like banking, health care and education.

Deployment models

There are four models in which cloud computing is deployed.

- Private cloud: Here the user is single organization. It is controlled and managed by IT department of the host organization. Eucalyptus system is the best example of private cloud.
- Public cloud: Here number of organizations can use it on shared basis. It may be hosted and managed by a third party acting as a cloud service provider. E.g Google app engine, Microsoft azure.
- Hybrid cloud: A single organization may opt to avail benefits accruing both from private cloud and public cloud. Such a use is called hybrid cloud. E.g. Amazon web services.
- Community cloud: Related or cognate organization may wish to make use of common cloud computing environment. This mode of use of cloud computing is termed as community cloud. E.g. Government cloud.

Three service models

- Software as a service [SaaS]: All applications availed as service typically through web browser by end user. E. g. salesforce.com, google apps.
- Platform as a service [PaaS]: Here a application development & deployment platform is provided as service to software developer, so that he can build, deploy application. E.g. Microsoft Azure, Google App Engine.
- Infrastructure as service [Iaas]: Here hardware resources are thrown open to end user for executing services using virtualization technology. E.g. Amazon S3.

2. ENERGY EFFICIENT DATA CENTER

Data center known to consume lot of electric power. Power generation, we know has more harmful effect on environment. So there is urgent need to make data center energy efficient. Problem of high power consumption can be tackled at number of stages (power conversion from ac to dc, server, memory and network). Decision regarding work consolidation and device state transition can be done through interaction between data center controller and resource specific controller for energy efficiency. While excising economy on power consumption by (data center) outmost care needs to be taken so that it never at the cost services provided to end user i.e. SLA violation must be kept as low as possible.

2.1 Server power management

Server power management is an area where proper control of power can yield good results. Here I discuss few power management policies that use the mechanism of dynamic voltage scaling and vary-on/vary-off.

2.1.1 Independent voltage scaling [IVS]: In IVS each node independently handles its own power consumption by making use of dynamic voltage scaling. This way IVS can easily account for energy saving of 20-30% [4].

2.1.2 Coordinated voltage scaling [CVS]: CVS uses dynamic voltage scaling in a synchronized way so that all nodes work extremely close mean frequency setting across data center. To put into practice processor must have software controlled dynamic voltage scaling. E.g. AMD'™ Power Now technology. CVS is amenable to better energy saving plans than IVS but not without some characteristic complexities unique to it.

2.1.2 Vary-On Vary-Off [VOVO] : it works like a governer and regulate the server nodes so that at a particular point of time only that much number of server kept active which are just sufficient to support the workload. This will require the introduction of some hardware support such as Wake On LAN interface to instruct a server to change from inactive to active state and vice-versa. So in VOVO it is the workload that determines and regulates energy saving.

2.1.4. Combined policy: It is union of VOVO and IVS. It gives more energy saving than either VOVO or IVS in isolation.

2.15 Coordinated Policy: in this policy the VOVO and CVS work in conjunction where VOVO restricts the use of active servers and CVS controls and reduce the power consumption of individual active nodes. VOVO-CVS combine help to save up to 18% more energy than VOVO alone but certainly not without complicated implementation.

2.2 Dynamic Thermal Management for data center

Another approach is by applying the concept of dynamic thermal management at the granularity of a complete data center rather than individual servers or chips. In this I want to discuss policies for *workload placement* that can facilitate uniform distribution of temperatures through active thermal zones. Data center are known to give their best at ambient temperature below 25^{o} C.

We have two approaches to using workload redistribution to achieve these thermal management goals and discuss how these add to enhance robustness and energy efficiency of cooling system.

A. Row-wise thermal management can determine these imbalances by redistributing workload locally within a row based on measured temperature, equalizing temperature variation within the hot aisles. Excess temperature rise of the ith rack of a row is defined as the difference between its drain temperature (Ti) and the temperature of the cold air entering the room (Tref)[5].

B. The regional thermal management method depends on the ability of a workload redistribution mechanism to move large compute loads around the PDC(Programmable data center) in the event of infrastructural problems (like cooling failure, power failure etc.) or a major increase in computational load.

This approach reduces energy consumption by more than 14% by workload placement.

2.3 Virtualization

When we observe that an application no longer need an exclusive server we can club number of application and can make them run on a single server off course within it’s capacity to handle that load. This is what we mean by term virtualization. It significantly decrease the amount of hardware required.

The virtual world products such as XEN, VMware and Hyper-Vare part of the green cloud computing, those virtual products have advanced methods to reduce the computing power consumption.[6] This virtual live migration of application from one server to another result in benefits such as improved performance, fault-tolerance and manageability, while allowing workload movement with a short service down time. There no denying the fact that service level of applications are slightly effected during a live migration. . Fig 2.3 shows energy saving through virtualization. Popular hypervisors, such as Xen and VMWare allow migrating an OS as it continues to run[7].

Fig 2.3 Energy saving through virtualization

Source: [Green Computing Conference, IEEE, 2010, pp. 357 – 364]

2.4 Live migration

Management of available physical resources is easily achieved in a flexible manner through live migration. It helps in balancing of load and maintenance of infrastructure but not at the cost application availability and responsiveness.

Memory transfer phases:

Push phase

The source VM remains in running mode while certain pages are pushed across the network to the new destination. To ensure consistency, pages modified throughout this process must be re-sent.

Stop-and-copy phase

The source VM is stopped, pages are copied across to the destination VM, then the new VM is started.

Pull phase

The new VM executes and, if it accesses a page that has not been copied yet, this page is faulted in ("pulled") across the network from the source VM[8].

2.5 Data center architecture

Huge **c**onsumption of power at data centers as always been a matter of concern for persons engaged in research & IT industry. Efforts are on to devise methods to reduce power consumption of data center which amounts to about 15.5% cost of the data center.

The DCell[9] data center architecture is created out of commodity mini-switches to scale-out. The architecture of DCell is recursive. The BCube data center architecture is designed to be applied in container based, modular data centers, which have a few thousands of servers. BCube is defined recursively, a BCubek is structured from n BCubek−1 and nk n-port switches. One of the design principles of fat-tree topologies is to build a data center network by making use of small, commodity switches. The fat-tree topology, also known as Clos topology, consists of three structural layers. Fig. 2.5 shows energy comparison of data architecture

In all cases, the power consumption of the architecture highly depends on the parameters of the architecture, namely the number of ports, denoted by n, that a server or switch can have and the number of structural levels, denoted by k.

Architecture	Power consumption	Diameter
Balanced tree	$En^k + E_{sw}\sum_{i=0}^{k-1} n^i$	2k
Fat-tree	$En^3/4 + E_{sw}\,[(n/2)^2 + n^2]$	6
DCell	$\approx (E + E_{sw}/n)(n+1)^{2k}$	$2^{k+1} - 1$
BCube	$En^{k+1} + E_{sw}\sum_{i=1}^{k+1} n^i$	k+1

Fig. 2.5 shows energy comparison of data architecture

Source: [Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, ACM, 2010, pp. 183-186]

2.6 Energy aware profiling

Huge energy demand of the data center necessitates tackling of the problem at different level in cloud computing to seek areas where consumption of power can curtailed without undermining service efficiency. Energy aware profiling helps us to better understand the way the energy is consumed and help the application developer from the application level enhance their decision-making in terms of energy-awareness when optimizing their applications and services. In real cloud environment play important roles are played by the physical desktop, virtual desktop and dynamic provisioning virtual desktop.[10]

This system architecture would have the RMU [Resource Monitoring Unit] to dynamically collect the energy devoted by the hardware components and observe the number of assigned VMs. After that, EPU [Energy Profiling Unit] have appropriate algorithms to calculate the energy consumed by each VM and hardware components, and it would then profile and populate these measurements as KPIs[Key performance Indicator] to a database. This data can be further examined by the Reporting and Analysis Unit to provide the software developers energy-aware reports in order to enhance their awareness of the energy consumption when making programming decisions.

Software developer can make best use of this new method to understand how their applications are using the infrastructure resources and consuming energy in order to improve their decision-making when creating and configuring new application[11].

3. CONCLUSION

Indeed cloud computing is a novel discrete way of delivering IT services to diverse users. And other remarkable thing about cloud computing is that it can be availed of in a metered way like other utility services. The only drawback with cloud computing is that it is a notorious power guzzlers and call for a stringent ‘energy efficiency’ regime. Here we may add a word of caution. The measures taken to reduce power consumption should in no way undermine the efficiency of Service Delivery. Like a tight rope dancer data center is in quandary to balance the two.

REFERENCES

[1] Dhingra A, Paul S, “Green Cloud Heuristic based BFO Technique to Optimize Resource Allocation”. Indian Journal of Science and Technology, 2014, pp. 685–691

[2] Kumari S.P., Kamal Banu Nadira A.R., “Optimal Integrity Policy for Encrypted Data in Secure Storage Using Cloud Computing”. Indian Journal of Science and Technology, 2016.

[3] Verma A., Sharma S., Kaur B.,”Comparative Analysis on Load Balancing Techniques in Cloud Computing”. Indian Journal of Science and Technology, 2016.

[4] Elnozahy ENM, Kistler M, Rajamony R, Energy-efficient server clusters. Power-Aware Computer Systems, Springer, 2002, pp. 179-197.

[5] Sharma RK, Bash CE, Patel CD, Balance of power: Dynamic thermal management for internet data centers. IEEE Internet Computing , IEEE, 2005, pp. 42-49

[6] Alzamil I, Djemame K, Armstrong D, Energy-Aware Profiling for Cloud Computing Environments. Electronic Notes in Theoretical Computer Science, 2015, pp. 91-108.

[7] Younge AJ, Laszewski G Von, Wang L, Efficient resource management for cloud computing environments, Green Computing Conference, IEEE, 2010, pp. 357-364.

[8] Clark C, Fraser K, Hand S, Hansen JG, Live migration of virtual machines. In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation , USENIX Association Berkeley, 2005, pp. 273-286

[9] Khajehei K., “Green Cloud and Virtual Machines Challenges”. Indian Journal of Science and Technology, 2016.

[10] Gyarmati L, Trinh TA, How can architecture help to reduce energy consumption in data center networking? In Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, ACM , 2010, pp. 183-186,.

[11] Muthu K., Somasundaram K., “Energy Efficient in Virtual Infrastructure and Green Cloud Computing: A Review”. Indian Journal of Science and Technology, 2014.