International Journal of Engineering & Technology, 7 (3.34) (2018) 141-144

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



A survey on Approaches Used for Efficient Workload Management in Geo-Distributed Data Centres

D.Ramya[1]*, J.Deepa[2], P.N.Karthikayan[3]

[1,2,3]Assistant Professor, Department of Information Technology
Veltech Rangarajan Dr. Sagunthala R&D Institute of Scienceand Technology, Avadi, Chennai – 600062.
**Corresponding author E-mail: dramya@veltech.edu.in*

Abstract

A geographically distributed Data center assures Globalization of data and also security for the organizations. The principles for Disaster recovery is also taken into consideration. The above aspects drive business opportunities to companies that own many sites and Cloud Infrastructures with multiple owners. The data centers store very critical and confidential documents that multiple organizations share in the cloud infrastructure. Previously different servers with different Operating systems and software applications were used. As it was difficult to maintain, Servers are consolidated which allows sharing of resources at low of cost maintenance [7]. The availability of documents should be increased and down time should be reduced. Thus workload management becomes a challenging among the data centers distributed geographically. In this paper we focus on different approaches used for workload management in Geo-distributed data centers. The algorithms used and also the challenges involved in different approaches are discussed

***Keywords*:** *Cloud Computing, Data Centers, Virtual Machine, Virtualization.*

1. Introduction

1.1. Cloud Computing

In older days the need for storage is very less. The data were mainly stored as documents with more paper works. Today the business grows bigger and bigger and so the organization contains more number of data need to be stored. The data are stored in digitized form and need to be handled carefully. Cloud computing is the technology used to handle large data stored globally[5].
Cloud computing provides computing services to the customers such as storage, databases, servers, software and networking through the Internet. The combination of hardware and software resources delivered as a network service is said to be Cloud computing. It is a virtual infrastructure which can be accessed on-demand and work on pay per use basis model. There are four types of cloud namely Public, Private, Community and hybrid cloud[1].
Cloud computing is classified based on two ways.
1) Cloud computing location
2) Types of services that cloud computing offered
Based on location the cloud computing is classified into three categories.

A) Public Cloud

In public cloud, the cloud vendors host the cloud infrastructure. The customer can store, process and retrieve the data but they can see the computing infrastructures physically. A public cloud is basically termed as an Internet. There are so many advantages the public cloud provides. It is low of cost as the cloud vendor takes over the infrastructure and maintenance. Scalability is very high. They provide the resources when the customer demands to meet the business requirements. It is reliable means that the system failure can be managed by other servers as the data are in multiple servers.

B) Private Cloud

The private cloud is created for the particular organization. It is not shared with other organization. Private clouds are more secure when compared to public cloud. It is more expensive and flexible.

C) Hybrid Cloud

The combination of both public and private cloud is said to be Hybrid cloud. The Organizations store critical applications on private cloud and common applications in Public cloud. The hybrid cloud ensures easy transformation of applications between public and private clouds. It is more flexible that the organization can use resources from public when they need.
Based on the types of services offered, the cloud computing are classified into three categories.
a) Infrastructure as a service (IaaS)
b) Platform as a Service (PaaS)
c) Software as a Service (SaaS)

A) Infrastructure as a Service

IaaS delivers the cloud computing infrastructure as a service, such as storage, servers, operating system and network. The organizations instead of buying those infrastructures separately they get them from the cloud as a service.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

b) Platform as a Service

The cloud provide software development platform as a service to the customers. This enables the customers to share the same platform among the different branches of the organization.

c) Software as a Service

This provides the customer software as a service. The software applications are stored in the cloud and users can access by paying for usage of such applications.

e) Benefits of Cloud computing

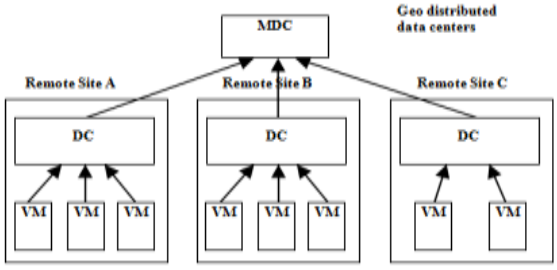
1) Reduction of cost to the customers
2) Speed operation
3) Availability of data at any time
4) Increase productivity of organizations
5) Improves performance and reduce latency
6) Reliable service

1.2 Terminologies used in Cloud Computing

A) Data Center

Data centers are locations with computing and networking equipment. It collects stores, distribute and process data. An organization can own the data center by itself maintained by trained employees. It can also be a storage center with collection of servers. The data can be obtained either by physically or virtually[3].

B) Data Center Architecture



The architecture of data center consists of Main Data center (MDC) and the distributed data centers (DC). The data centers from different locations connect to the main data center. The data centers have so many virtual machines connected to it. Virtualization enables the customer to access the information at any time from any locations.

C) Virtualization

Virtualization is defined as the creation of virtual version of a server, storage device, network, desktop and operating system. There are so many types of virtualization which includes the following: Server virtualization, Storage virtualization, network virtualization, Application virtualization and Operating system virtualization.

D) Virtual Machine

Virtual Machine (VM) is an operating system or software program that reveals the behavior of the computer. It is capable of working like a separate computer. VM can offer Instruction set architecture which is totally different from real computers. The maintenance of VM is very easy. It ensures availability of data to the customers at any time. Data recovery becomes efficient in VM as they have multiple copies. Multiple OS environments can be available in the same machine but confined in different environments.

2. Comparative Study of Different Approaches

2.1 Temperature Aware Workload Management

A generalized m-block ADMM (Alternating Direction Method Of Multipliers*)* algorithm inspired by [20, 23]. Convergence theorem is given for the algorithm[10]. Based on the m-block ADMM algorithm, a distributed solution algorithm was developed.
Two key issues were considered.

A) Cooling systems

I) Data Center-Cooling
Air conditioner in the computer room contains CRACs (computer room air conditioner). The cooling coil present in CRACs absorbs hot air exhausted from server. The water in the cooling coil extracts the heat. The returned hot water from CRACs is cooled through compressor of the mechanical chiller. This compressor accounts more cost and also consumes more energy of data center power[6].

Ii) Outside Air Cooling
In order to improve the efficiency of energy, free cooling technologies are used which works without mechanical chillers. The technology adopted called outside air cooling. Air-side economizer is used in CRAC to push the outside air in to datacenters for cooling purpose. The hot air exhausted is expelled out. It is not recycled and used. This technology totally depends on outside temperature. If the temperature is low, less air is needed for cooling the data centers. The idea is to manage the workload temperature aware by using outside air cooling.

B) Request Routing and Batch Workloads

The energy consumption takes place in both interactive workloads and batch workloads issued by the user. If a certain location is cost efficiency, the capacity allocation can be increased for interactive workloads instead of batch workloads. Dynamic capacity allocation is done by request routing.
The ADMM algorithm and Distributed algorithm are used for request routing and capacity allocation. ADMM is very simple and also very powerful algorithm. It overcomes the drawbacks of previously used dual decomposition methods. It gained practical use in many problems. The distributed algorithm can be implemented parallel in many servers.

2.2 Heterogeneity-Aware Workload Management

The algorithm of placement and the algorithm of migration were used for Heterogeneity-aware workload Management.
The approach develops heterogeneity aware algorithms. sCloud developed to implement these algorithms to maximize the output of the distributed data centers. The main points considering the sCloud is allocation of cloud workload and the distribution of batch jobs to different data centers [2].
The constraint considered for system good throughputs are green power supply that varies depends on the time, characteristics of heterogeneous workload and QoS requirements. The placement algorithm which is based on optimization used to place the transactional request to the distributed datacenters. Another online algorithm is integrated to dynamically migrate the batch jobs to other data centers based on green power supplies. This algorithm used to improve the system output in higher range.

Architecture of sCloud

1. sCloud collects weather forecast information at different geographical location and assigns the green power.
2. sCloud detects the workload intensity and the application's performance in each data center.
3. The decision should be taken on how to allocate resources and where to place the transactional requests and when to migrate the jobs across data centers.

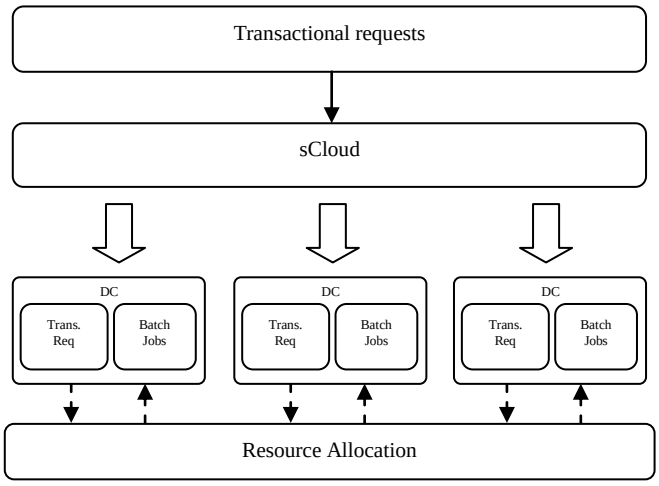


Fig1: Architecture of sCloud

Description

A) Prediction of Green Power

Due to varying weather conditions, the solar and wind energy generation may not be in a consistent range. Prediction methods are used to determine the amount of green power in a particular period of time. Workload placement and allocation of resources in distributed data centers are done based on the prediction.[8]

B) Placement Algorithm

sCloud considers input such as current weather forecast, arrival rate of workload and the rate of service. The workload is placed on the data centers and resources are allocated evenly based on those inputs to achieve the good output.

C) Job Migration Algorithm

Though the placement algorithm manages to distribute the workload, the job is migrated to other data centers based on the availability of power. This helps to maximize the system output and to meet QoS requirements. Job migration algorithm determines the job to be migrated and the data center to which it should be migrated. sCloud follows two steps for job migration.
First, select the particular data center among the distributed data centers. If particular job is to be migrated, the job completion time of current data center and other possible data centers are calculated at a particular interval.
Secondly, the job migration decision is made based on the comparison of performance between current data center and selected data center. If the completion time of current data center is longer, then the job is migrated towards the selected data center. Otherwise the current data center completes the job.

2.3 Hierarchical Approach for Workload Management

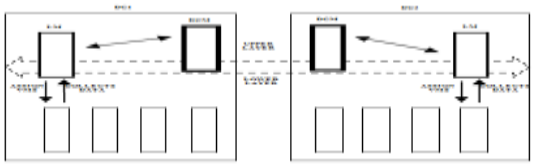


Fig2: Sample EcoMultiCloud Framework with two interconnected data centers

EcoMultiCloud uses hierarchical architecture[9] with two layers.

Upper Layer

The upper layer determines the distribution of VMs(Virtual Machines) between the DC's(Data Centers) and also exchange information among different DCs. The DCM (Data Center Manager) performs the functions of upper layer. It runs the algorithm and take decisions for migrating an workload from one site to another. The algorithm of upper layer can be changed without affecting the operation of the site.
DCM collects the information from lower layer and do the functionalities of the upper layer. DCM acquires detailed knowledge of the running VMs, the resource usage of the VMs and state of the DC. After gathering this information, it transmits to other DCs. DCM deals with high level information of a single DC such as energy consumption, energy costs, carbon emissions and Quality of Service. It summarizes the information and transmits to the other DCMs which in turn is very useful for assignment and migration of jobs [4].

Lower Layer

The lower layer allocates the workload inside the single DC's. Each DC in this layer is totally autonomous. They can manage their workload by using algorithms. The LM (Local Manager) takes care of the lower layer functions.
Assignment algorithm and Migration algorithm were used for this approach

A) Assignment Algorithm

The distribution of workload is defined by the assignment function. The assignment function fassign is defined as follows.

$$f_{assign} = \alpha.\frac{Ei}{Emax} + \beta.\frac{Ui}{Umax} + \gamma.\frac{Ci}{Cmax}$$

Where E, U and C represents Carbon emissions, Overall utilization and Costs of energy respectively. The positive coefficients are α, β, γ and their summation is 1.
The value of f_{assign} is computed for each DC, the VM is assigned to DC whose value is very low. The assignment of VM to the particular physical host in the corresponding DC is done by local assignment algorithm. The three terms are calculated with the small values that DCM transmits to the others.

B) Migration Algorithm

The assignment algorithm used for distribution of VMs becomes inefficient when the price or load of the energy varies in one or more of the data centers. The f_{assign} values of two DCs are compared with the fixed threshold value. If the value varies, then migration algorithm is used to redistribute the workload across the DCs.

3. Conclusion

Cloud computing has become the most needed technology nowadays. The workload management plays a vital role in cloud com-

puting to ensure the proper distribution of jobs among the data centers and also across the data centers. In this survey, we studied the three approaches used to manage the workload. There were different algorithms used for distribution of jobs efficiently. The management of workload increase the overall output of the system.

References

[1] Various green datacenters. http://www.ecobusinesslinks.com/.

[2] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," Future Generation Computer Systems, vol. 28, no. 5, pp. 755–768, 2012.

[3] F. Kong and X. Liu, "A survey on green-energy-aware power management for datacenters," ACM Comput. Surv., vol. 47, no. 2, pp. 30:1–30:38, Nov. 2014.

[4] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser. "Renewable and cooling aware workload management for sustainable data centers". In Proc. ACM SIGMETRICS, 2012.

[5] S. Ren and Y. He. Coca: Online distributed resource management for cost minimization and carbon neutrality in data centers. In Proc. SC, 2013.

[6] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser. Renewable and cooling aware workload management for sustainable data centers. In Proc. ACM SIGMETRICS, 2012.

[7] C. Li, R. Zhou, and T. Li. Enabling distributed generation powered sustainable high-performance data center. In Proc. IEEE HPCA, 2013.

[8] D. Gmach, J. Rolia, C. Bash, Y. Chen, T. Christian, A. Shah, R. Sharma, and Z. Wang. Capacity planning and power management to exploit sustainable energy. In Proc. IEEE CNSM, 2010.

[9] A. Forestiero, C. Mastroianni, M. Meo, G. Papuzzo, and M. Sheikhalishahi, "Hierarchical approach for green workload management in distributed data centers," in 20th International European Conference on Parallel and Distributed Computing, Euro-Par 2014, ser. LNCS, vol.8805. Porto, Portugal: Springer, August 2014, pp. 323–334.

[10] XU, H., FENG, C., AND LI, B. Temperature aware workload management in geo-distributed datacenters. Tech. rep., University of Toronto.