International Journal of Pure and Applied Mathematics
Volume 117 No. 21 2017, 219-229
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

A Virtual Machine Based Load Balancing Algorithm for Cloud Computing

[1]Venkateshwarlu Velde and [2]B. Rama
[1]Department of Computer Science,
Kakatiya University,
Warangal, TS.
veldevenkat@gmail.com
[2]Department of Computer Science,
Kakatiya University,
Warangal, TS.
rama.abbidi@gmail.com

Abstract

The rapid development of cloud computing, bring great convenience to developers. Recently the resource management of the cloud platform has become a hot research topic, especially the load balancing problem in data center is very important for cloud provider. In this paper, the challenges i load balancing are discussed load balancing framework is proposed for cloud platform, it use the threshold window strategy and an advanced AR prediction model to reduce the migration of VMs. Algorithm proposes that this method can effectively achieve load balancing, promote the utilization of the physical machines, and solve the frequent migration problem caused by high instantaneous peak values significantly.

Key Words: Cloud computing, load balancing, threshold window.

1. Introduction

Cloud computing is primarily based on virtualization era, thru the community services to provide customers with basic assets, utility systems, software and different services [1]. In the case of IaaS (Infrastructure-as-a-Service), operators provide a whole software application environment by way of offering hardware, software program and machines close to the service-first-class agreements reached with customers. VM (VirtualMachine) era is often utilized in information facilities, cluster computing and other packages. The era lets in a couple of working structures can run on the identical operating system, and provides a reliable isolation, significantly enhancing the reusability of physical sources. However because of the uncertainty of the useful resource usage of the user software, a trouble of low configuration or over configuration happens [2].

VM migration method can maximize infrastructure and balance facts center workloads with out violating Service Level Agreements (SLAs) [3]. Some of the prevailing VM migration methods are static algorithms which select the fine node for migration .These techniques are based on the present execution and hardware workload facts [4].The dual-threshold approach [5] is a traditional static VM migration that units both the higher and lower bounds of the host's workload and initiates the migration whilst the workload exceeds or falls underneath the decrease bound. However, the edge-based totally migration strategy lacks the potential to expect destiny workload tendencies for the host device, and hence may additionally trigger unnecessary and wasteful VM migration if the workload of the host is most effective a temporary top. Dynamic techniques commonly depend on existing reputation and historic workload records to make decisions. Beloglazov et al. [6] proposed an adaptive energy performance and threshold based heuristic algorithm, with the aid of tracking the resource utilization to control VM migration. Zhao and Shen [7] used the autoregressive model of time collection prediction generation to expect destiny values and are expecting future load values based on future time collection. Mao et al. [8] proposed an automated scaling mechanism primarily based on load records to automatically scale cloud times. The Nomad BIOS [9] is a virtualized migration machine based totally at the L4 microkernel. It makes use of pre-replica migration to acquire the downtime as small as possible. Mong[10] emphasised the capacity software of multi-agent device inside the automatic cloud carrier composition, the complex negotiation process of cloud commerce and the hunt of cloud offerings. Cao et al. Use multi-agent system generation to manipulate cloud services, making sure the best of service [12].

This paper employs an advanced load balancing framework for information middle, the use of dispensed approach primarily based on time collection workload prediction and threshold window strategy to make decisions in VM migration.

2. Challenges & Issues of Load Balancing

Even via cloud computing is extensively used in recent times however nonetheless the research is in its strategy planning stage. So before explaining the weight balancing algorithm for cloud computing it is required to perceive a few key challenges and problems that have an effect on the overall performance of load balancing algorithms. The Automated provider provisioning – Elasticity is a main component in cloud computing due to which allocating and releasing of the assets take place as default. The task with using most excellent aid is how cloud elasticity may be used and how work with conventional structures overall performance can be carried out concurrently

Virtual Machine migration – The concept is to assume a machine as a set of documents or a document. It is feasible to decrease the burden on over loaded machine by way of transferring the virtual machine amongst them in powerful manner. The principal motto is to distribute the all type of load in a datacenter. The mission is to do away with and avoid drawbacks of cloud computing device when the burden is dynamically distributed by using virtual machine

Energy management – Energy Management is likewise a fundamental factor that allows customers to use the sources from global centre. This gives financial system of scale and predominant advantage to cloud computing however a question emerges that via the use of just a portion of data enter, the way to meet better performance.

Stored records control – Another key requirement is the storage of records. So how can records be disbursed within the cloud system with maximum appropriate memory and rapid access?Vurukonda, N et al [11] Time Based Encryption is an open key based encryption that draws in gets the opportunity to control over encoded Information utilizing access approaches and credited properties.

Emergence of small distinct data enters for cloud computing – A small length of data centre can be more useful just because it will devour much less electricity and cheaper than large one. And load balancing is showing as a worldwide scale problem for certifying right response time with foremost aid usage and distribution.

Spatial distribution of the cloud nodes – Some algorithms are proposed just for almost placed nodes in which communiqué delays are insignificant. But nonetheless it's far a problem to layout an efficient load balancing algorithm which can be formulated nicely for spatially dispensed nodes.

Storage and Replication – A full replication algorithm is not a good deal useful for efficient storage usage in a device. This is just due to the fact the same amount of records will be stored in all replicated nodes. Full replication algorithms motive unreasonable costs with requirements of massive garage.

Algorithm complexity – Load balancing algorithm is preferred to be much less complicated in phrases of operations and execution (implementation). A poor implementation complexity will result in a more complicated manner. Furthermore, for monitoring and controlling the implementation, algorithms require better communiqué, more records and delays may also cause more bottlenecks and then Performance discards.

Point of failure controlling – Some algorithms (centralized algorithms) provide effective mechanisms for processing load balancing in a specific pattern. But the difficulty is that there's best one controller for the whole system. In such circumstance, if the controller fails, then the entire machine fails.

3. Load Balancing Framework for Cloud Environment

The green resource allocation algorithm proposed via Rajkumar Buyya et al [6] units up a monitoring thing on each physical machine to reap the workload records, then the information is transmitted to the VM Manager (Virtual Machine Manager), which receives and stores the ancient workload information and makes the prediction.

Load Balancing Framework for Data Centre

Figure 1 indicates the cloud model based totally on inexperienced useful resource allocation version. There are basically numerous entities involved

1) Users/Brokers

Cloud customers or their agents can post carrier requests to cloud from everywhere within the global. A broker is an individual or company acting as an middleman between the user and dealer in the cloud computing services.

2) Other components

According to Rajkumar Buyya et al[6], there are other components in Green Resource Allocator: Green Negotiator, Service Analyser, Consumer Profiler, Pricing, Energy Monitor, Service Scheduler, Accounting.

3) VMs

VMs can be dynamically started out and stopped on a single physical system to meet established requests, for this reason offering maximum flexibility to configure numerous partitions of assets at the same physical system to distinctive specific requirements of provider requests. Multiple VMs also can concurrently run programs based totally on specific running machine environments on a single physical machine. In addition, with the aid of dynamically migrating VMs across physical machines, workloads can be consolidated and unused assets can be positioned on a low-power state grew to become off or configured to function at low-overall performance degrees (e.g., the usage of DVFS) so that it will keep energy.

4) Physical Machines

The underlying physical computing servers offer hardware infrastructure for developing virtualized sources to meet service demands.

5) Workload Agent

This element is built in every physical machine. It collects the information of the physical machine and the VMs and keep the historic workload information.

6) VM Manager

Manage virtual machines on physical machines. Collect the workload information from every agent on physical machine. Making choice of the virtual machine migration, using our strategy to migrate the virtual machine. It's the key factor inside the load balancing work.

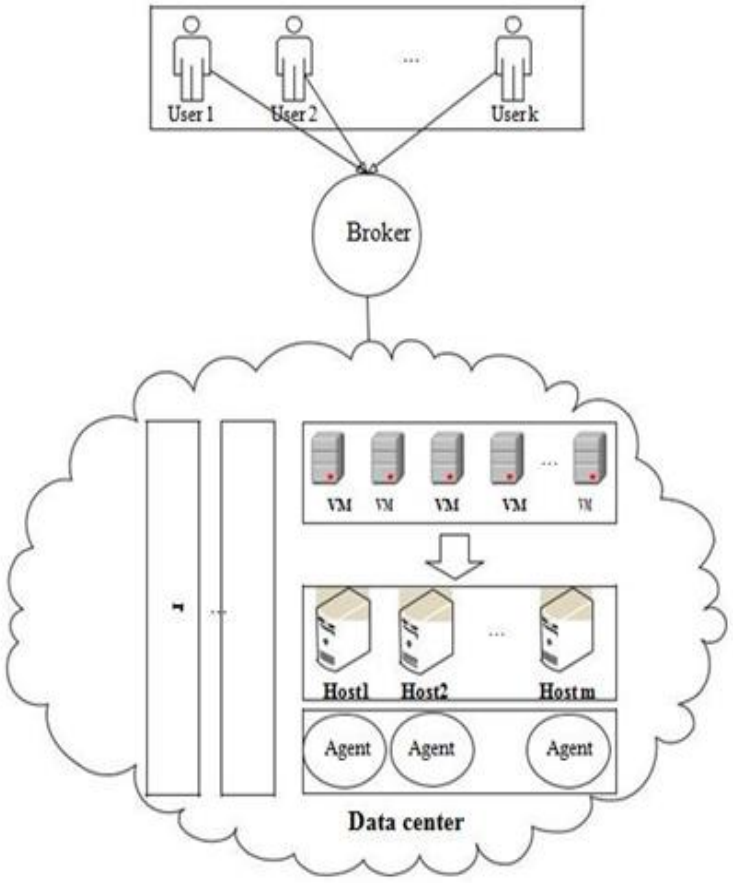

Figure 1: A Load Balancing Framework

VM Migration Architecture

All the VM migration is determined through the VM supervisor. The controlling concept of our VM migration method can be described as workload-aware migration (WAM). And the weight balancing process is particularly shown in Figure 2.

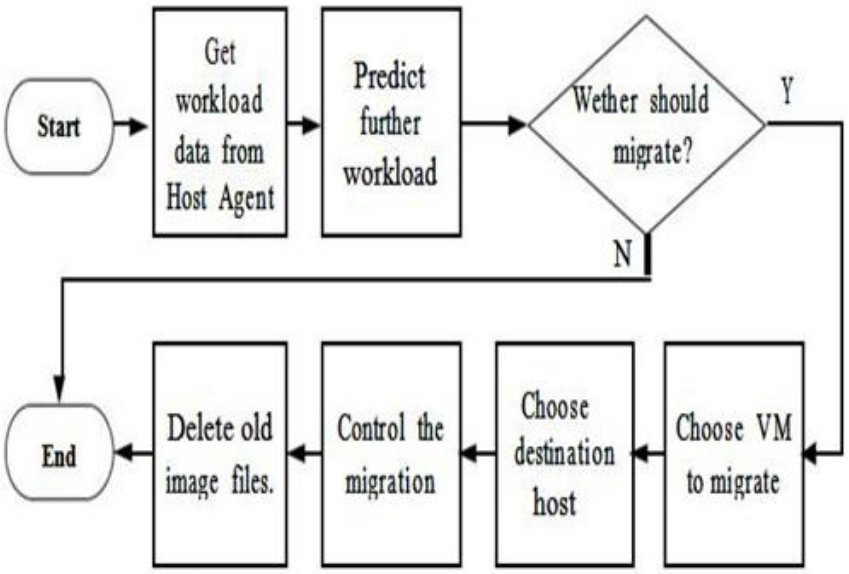

Figure 2: Migration Process in VM Manager

Workload agent collects the workload facts and transports it to VM manager in real time. VM manager predicts destiny workload, determines whether or not to migrate the VM or now not, and determines which VM can be migrated, then chooses the appropriate destination host. Finally VM Manager implements the migration, and deletes the replicate photograph documents associated with the removed VM at the supply host device. In addition, VM Manager considers the response time, migration frequency and different troubles in decision-making.

4. Load Balancing Migration Strategy based on Agent

The particular algorithms and techniques used on VM supervisor or migration technique can be delivered in this section.

Advanced Time Series Workload Prediction

The workload records of a physical machine are fashioned with the aid of its real beyond workload values which can be accrued by way of the facts series module at the workload agent. A device's CPU workload is used as the indicator for the whole workload of that machine due to the fact that a better CPU usage rate shows a higher consumption of aid of the device. The workload agent on each host collects workload statistics as soon as for the duration of a time which is named as T. Due to the uncertainty of the workload, workload facts that is too old has no impact for prediction. So, the agent most effective saves the today's N workload sequences, which can be carried out with a queue. Because of the uncertainty of the workload at the physical machine, AR(Auto Regressive) technique is hired for our time-series prediction trouble, that could

give a better reaction time and better effect in small records prediction than ARMA(Auto Regressive and Moving Average) and ARIMA(Auto Regressive Integrated Moving Average). AR is linear predicting models that may are expecting the former or subsequent of the acknowledged-facts. In this paper, AR model with N order (additionally known as AR (N)) is used, the following expect value xf(n+1) is determined through the preceding N-period time collection workload collection values of xn and a deviation term. The method is as follows:

$$x_{f(n+1)} = \sum_{n=0}^{N} a_n x_{t-n} + e_t$$

[1]Where in $\{a_n\}$ is the version parameters, and the et is a white noise deviation time period with an average price of 0, that are determined through model training. In actual use, we are expecting the future several values, each with ultra-modern history workload and then take their average because the predicted fee, so that we can get more uniform prediction consequences. We use N time collection values to expect the destiny M workload (M<<N), after which take the common Wf as the destiny predictive value of the AR version. Its definition is as follows [2]

$$W_f = \frac{\sum_{m=0}^{M} x_{f(N+m)} + \sum_{m=0}^{M} x_{N-M}}{2 * M}$$

Wherein N is the number of historic workload information, M is the number of workload cost we forecast, xf(N+M) is workload cost or predictive value, xN-M is the history workload.

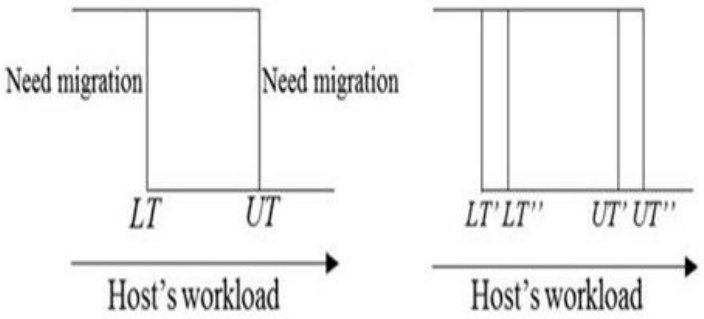

Figure 3: (a) Thresholds of Traditional Methods; (b) Threshold Windows of our Method Based on Workload Prediction

Whether to Migrate

Traditional static approach sets two thresholds to decide whether a VM is wanted to migrate or not, as shown in Figure 3(a). When the workload is more than UT or decrease than LT, We need to migrate out some VMs.

Choose VM to Migrate

The decision principle is: pick out a VM that has not been migrated recently and the VM workload is quite big. The service nice can be affected, even violate SLA, at the same time as a VM is migrated regularly. So a time threshold Tth is about to keep away from this situation. Each VM facts a time fee Tmigrate which way the final time being migrated to this physical machine (if no longer migrated ever, take the creation time as a substitute). So the approach is given out as: if a VM turned into migrated in Tth, mark the VM can't be migrated now, then pick the migrate VM which occupies the biggest workload and may be migrated. After migration, if the workload continues to be too excessive, maintain to migrate once more.

Choose the Vacation Spot Host

According to our strategy, the principle elements are: the machine can accept the workload of VM which is migrated; the migrated VM can improve the usage ratio of this host's assets. Therefore the precept is to enhance the CPU usage charge while not exceeding the top threshold restriction. So the approach is given out as: calculate a rating (commonly CPU idle rate may be used) for every physical device, and select the host because the destination host which get the very best score (if there are numerous highest scores, choose any one is OK).

5. Conclusion

Cloud computing is a new paradigm wherein distinctive sources are access by using a couple of customers over the internet in on call for demand. These sources are unexpectedly growing and also growing uses of heterogeneous machine in dynamic environment. But there are numerous studies challenges in cloud computing. Load balancing is important mission (issue) in cloud computing.

The key intention of load balancing is to satisfy consumer's needs via distributing work load among more than one node in device and maximize resource usage and improves machine performance. So efficient load balancing is critical for machine overall performance, useful resource utilization, stability, maximizes the throughput and minimizes the response time that are the principal objectives of this paper. To balance the load among a couple of nodes in system, there are several load balancing algorithms will be delivered.

A load balancing scheme is proposed in this paper, which coupled with the brink window method and superior AR prediction model. Our proposed method assumes that our strategy now not only remedy the common migration problem, however also advanced the reliability of the statistics center. In future, we would really like to explore other troubles in this framework along with distributed response, fast reaction time, and error dealing with.

References

[1] Xing Y., Zhan Y., Virtualization and Cloud Computing [M]// Future Wireless Networks and Information Systems, Springer Berlin Heidelberg, (2012).

[2] Chaisiri S., Lee B.S., Niyato D., Optimal virtual machine placement across multiple cloud providers[C], IEEE Asia-Pacific Conference on Services Computing APSCC (2009), 103-110.

[3] Garg S.K., Gopalaiyengar S.K., Buyya R., SLA-Based Resource Provisioning for Heterogeneous Workloads in a Virtualized Cloud Datacenter [M], Algorithms and Architectures for Parallel Processing, Springer Berlin Heidelberg (2011), 371-384.

[4] Schroeder T., Goddard S., Ramamurthy B., Scalable Web server clustering technologies [J], IEEE Network the Magazine of Global Internetworking14(3) (2000), 38-45.

[5] Chang Y.C., Chang R.S., Chuang F.W., A Predictive Method for Workload Forecasting in the Cloud Environment [M], Advanced Technologies, Embedded and Multimedia for Human-centric Computing, Springer Netherlands (2014), 577-585.

[6] Beloglazov A., Buyya R., Energy Efficient Resource Management in Virtualized Cloud Data Centers[C], IEEE/ACM 10th International Conference on Cluster, Cloud and Grid Computing, Computer Society (2010), 826-831.

[7] Zhao Y., Shen L., Application of time series auto regressive model in price forecast, IEEE International Conference on Business Management and Electronic Information (BMEI) (2011), 768-771.

[8] Mao M., Li J., Humphrey M., Cloud auto-scaling with deadline and budget constraints, IEEE/ACM 11th International Conference on Grid Computing (GRID), (2010), 41-48.

[9] Hansen J.G., Henriksen A.K., Nomadic operating systems (Doctoral dissertation, Master's thesis, Dept. of Computer Science, University of Copenhagen, Denmark), (2002).

[10] Vishnu S., Vignesh S., Surendar A., Design and implementation of ZETA micro-inverter for solar PV application, International Journal of Mechanical and Production Engineering Research and Development 7 (4) (2017), 215-222.

[11] Lakshmi K., Surendar A., Verification of axiprotocol using system Verilog, International Journal of Mechanical Engineering and Technology 8 (5) (2017), 588-595.

[12] Vurukonda, Naresh, Thirumala Rao B., Secure Sharing of Outsourced Data in Cloud Computing with Comparison of Different Attribute Based Encryption Schemes: A Review (2017), 680-698.

[13] Can B.Q., Li B., Xia Q.M., A service-oriented QoS-assured and multi-agent cloud computing architecture, Proc. of the Cloud Computing, Springer-Verlag (2009), 644-649.