**1. INTRODUCTION**

Cloud Computing is a delivery of shared computing resources over internet. It is defined as a type of computing that relies on sharing computing resources rather than having local servers or personal [devices](http://www.webopedia.com/TERM/D/device.html) to handle [applications](http://www.webopedia.com/TERM/A/application.html). It is used for storing and accessing of data and programs. Cloud computing enables users to consume resources as per their utility rather than having to build and maintain computing infrastructures in-house.  There are three major benefits of Cloud Computing. First is the benefit of [Self-service provisioning](http://searchcloudprovider.techtarget.com/definition/User-self-provisioning) which is servicing the end users with resources [on-demand](http://searchdatacenter.techtarget.com/definition/on-demand-computing). Secondly [Elasticity](http://searchcio.techtarget.com/definition/IT-elasticity), Users can scale up as computing needs increase and then scale down again as demands decrease. Third is [Pay per use](http://searchcio.techtarget.com/definition/metered-services), Cloud allows users to pay only for the resources and workloads they use. Cloud computing provides resources at door step. In other words, with internet facility Cloud computing can be done anywhere and can be accessed from any place.

After the evolution Cloud Computing, Virtualization comes into picture. Virtualization plays a major role in this fast developing environment. It is a software which makes it possible to run multiple Operating Systems and multiple applications on the same server at the same time. The essential component in the virtualization stack is the hypervisor. This hypervisor, also called as Virtual Machine Monitor (VMM), creates a virtual platform on the host computer.On top of this Virtual Machine Monitor, multiple guest operating systems are executed and observed. By this way, multiple operating systems, which are either multiple instances of the same operating system ormultiple instances of different operating systems, can share the hardware resources offered by the host.

Though virtualization benefits the users, the resource utilization levels increases drastically. Due to the changes in the resource requirements of the users, some Physical Machines may be heavily loaded while the others may be lightly loaded. This is decided based on the CPU utilization of the system. Actual CPU utilization varies depending on the amount and type of managed computing tasks. Certain tasks require heavy CPU time, while others require less CPU time because of non-CPU resource requirements.

When the Physical Machine is lightly loaded it indicates that the user underutilizes the requested resources. On the other hand, when the Physical Machine is overloaded, Quality of Service will be decreased. This overloaded condition of a Physical Machine is known as HOTSPOT. In other words, Hotspot can also be defined as the state when the performance of a system falls below the minimum acceptance level. During the occurrence of Hotspot, the downtime of the Physical Machine will decrease. That is, the performance of the system may go down. As the performance of the system decreases,

For this project, virtual machine is created using XEN as a hypervisor. Hypervisor acts as an abstraction layer above which any number of virtual machines can be created. Hypervisor makes it possible for multiple guest operating systems to run on a single computer.When multiple Virtual Machines run on a single Physical Machine, there is a chance of occurrence Hotspot, which is an overloaded condition of the Physical Machine. This Hotspot is detected via an IQR algorithm. CPU utilization of all the Virtual Machines are taken and the Threshold value is calculated using this algorithm. The output Threshold value is then compared with the CPU utilization of each Virtual Machine in the Physical Machine. As a result, the overloaded and the under-utilized virtual machines are determined. Now the overloaded Virtual Machine is migrated to the under-utilized Physical Machine. This is Virtual Machine migration is done to sustain the system performance. A Time Series based Analysis is made to maintain the system’s downtime during the time of virtual machine migration.

1. **LITERATURE REVIEW**

**2.1. CORRELATION ANALYSIS:**

The correlation analysis approach accurately identifies the bottleneck resource in the running state of the VM-hosted application with a little information. It quantities the correlation degree between workload parameters and resource requirement and overcomes the time lag by utilizing traditional monitoring approaches. The correlation analysis approach determines the bottleneck resource of a given application. In this section, based on the grey system theory, the WCM constructs a quantitative relationship between the cost of the bottleneck resource and the application workload by modeling the resource consumption. As the interval of the end-user, access requests is much shorter than the sampling interval of VMM monitoring. Virtual Machine Monitor cannot obtain a large amount of monitoring data on the measured resource consumption. Therefore, some traditional analysis methods, such as regression and multiple correlations, cannot be used to model the workload of a virtual environment. In this section, we use the grey system theory for the workload modeling task. This method is used because it can capture the resource consumption behavior in small scale using feedback mechanism and also avoids large-scale data processing. The grey system theory is more suitable for workload modeling in a virtual environment with a small amount of measured data. Hence for detecting Hotspot grey system theory of correlation analysis is used.

**2.2. MAP REDUCE:**

Hadoop provides a distributed file system (HDFS) and a framework for the analysis and transformation of very large data sets using the Map Reduce paradigm. One of the most important characteristics of Hadoop is partitioning and processing data in parallel. The HDFS is implemented as a user-level file system, providing the global access to files in a cluster. The files stored in a cluster are divided into many blocks, each of which isstored as a separate file in HDFS .It contains two services: Name Node and Data Node. Name Node is responsible for maintaining metadata of all files. It also takes care of the mapping between metadata and blocks of files. Whereas Data Node is the reallocation where all the blocks of files are stored. Using the Map Reduce programming model to deal with a large volume of monitoring data, we go deep into the programming model of Map Reduce, in order to take full advantage of its strong parallelism. In this mechanism, there are two components to detect hotspot. They are Mond and workload hotspot detector. Mond is a daemon which monitors the resource usage of each node and their virtual machine. Each node consists of group of virtual machines. Workload hotspot detector computes the workload of each node in a virtual machine and detects hotspot there exists. Mond daemon monitors resource usage of the nodes and their virtual machines at a periodic intervals. It also monitors the nodes and their virtual machines and sends the report to sink file. Here the CPU pressure gets decreased. This mechanism includes two other components. They are the Mapper and the Reducer. Mapper part considers all the data stored in the sink file, and it computes the workload of all the nodes and their virtual machines. This output of the Mapper part is sent as the input to the Reducer part. After the reader inputs the workload data, the reducer begins the process of detecting the hotspot. Finally the Reducer generates the hotspot there exist.

**2.3. VIRTUAL MACHINE SELECTION POLICY:**

Virtual Machine selection is done to identify the correct Physical Machine to which the virtual machine with hotspot should be migrated. Hotspot is the overloaded condition of the system. That is, the system is unable to service with the resource demand of the virtual machine. Hence the overloaded virtual machine should be migrated from one physical machine to another physical machine. This selection is done with the help of two virtual machine selection policies. These policies gives an idea about which virtual machine should be migrated and where it should be migrated.

**a) MedianMT (Median Migration Time):** In this algorithm, the migration time of all the virtual machines are calculated and the median time is computed. Out of all the virtual machines, the virtual machine with the median migration time is selected for migration. If there occurs a tie among virtual machines, a random selection is done. The major advantage of this algorithm is that it provides less energy consumption. On the other hand, its drawbacks are it fails to resolve hotspot many a time; Number of migrations required is very high; It violates Service Level Agreement; And its performance is also less. Hence, we go for MaxU algorithm.

**b) MaxU (Maximum Utilization):** In this algorithm, the virtual machine with the highest possible CPU usage is chosen to migrate from overloaded machine to another. That is the virtual machine which has the maximum CPU utilization value is chosen to migrate. This algorithm reduces the number of migrations and the resource energy consumption. Also it provides better performance on comparison with MedianMT. Hence this algorithm is the best solution to avoid hotspot.

The other approaches for virtual machine selection are as follows:

**(i) Resource constrained virtual machine:** The virtual machine, whose resource requirement cannot be locally fulfilled, is selected for migration. During the occurrence of hotspot, this approach is an easy way to find the virtual machine.

**(ii) Holistic approach**: At the time of hotspot occurrence, it may not always be efficient to select the overloaded virtual machine for migration. This is because if the virtual machine, facing a resource deficiency, is utilizing a large amount of memory then the time and effort required for migration will be high. Hence a comparatively smaller memory virtual machine is selected for migration. This is done because the time and effort required for migration will be less. Hence the memory freed by the smaller memory virtual machine can then be allocated to the larger memory virtual machine.

**(iii) Affinity based**: This approach is for effective communication of two virtual machines that is located in two different Physical Machines. If two virtual machines are communicating with each other, it is better to place them on the same physical machine. This will reduce the overall communication cost among the virtual machines by reducing network usage. On other hand, memory is shared between the two virtual machines. This affects the virtual machine selection during the time of migration. Migrating a virtual machine to a physical machine where the virtual machine which shares memory with the migrated virtual machine resides results in effective memory usage. Virtual machines, which share memory, can be migrated together with less effort. This is because similar content memory pages are used by the two virtual machines. Hence the memory pages transferred only once is sufficient. Such a scheme is known as gang scheduling of virtual machines.

**2.4. MIGRATION STRATEGY:**

There are two strategies to migrate an overloaded virtual machine from one physical machine to another. These strategies include Push and a Pull Strategy.

**1) Push Strategy:** In the case of the push strategy, an overloaded PM plays the role of the seller. The PM multicasts information about these items to the buyers. This information contains the current load and memory usage of each VMs. The buyers may bid on one, several, or all VMs. A bid consists of the PM’s available resources as well as of the PM’s past resource usage. The seller selects the best candidate from the bids and performs the migration. Only one out of all bids will get selected and a single VM is migrated. The push strategy fails if none of the buyers has sufficient resources to hold the seller’s VMs.

**2) Pull Strategy:** On the other hand when the pull strategy is used, the seller is an underutilized PM that advertises free resources. The other PMs in the system are potential resource buyers that bid with the characteristics of their VMs as described above. The seller decides which bid it accepts and initiates the VM migration.

A Physical Machine is considered overloaded if its load exceeds the threshold value. However, a migration from one Physical Machine to another should not overload the destination host system. Therefore, the auction will fail. A successful push scenario is by migrating a Virtual Machine to a Physical Machine so that the load on both Physical Machines will decrease below the threshold and the system will attain a more balanced state. In this case, when the load on a Physical Machine drops below the threshold value, the host considers itself under utilized and initiates an auction to sell its free resources. The ultimate goal is to move workload without causing the buyer’s load to drop below the threshold.

**3. PROBLEM DEFINITION:**

**3.1. PROBLEM STATEMENT:**

To maintain the performance of the Physical Machine, to sustain the system’s downtime and for the effective usage of resources during the occurrence of Hotspot without affecting the user interaction.

**3.2. PROBLEM DESCRIPTION:**

The resource utilization of Users increase drastically day by day. As the users demands keep changing, the load of the Physical Machine cannot be fixed. This changing demands leads to variation in the resource utilization levels of different Physical Machines. Due to the changes in the resource requirements, some Physical Machines may be heavily loaded while the others may be lightly loaded. When the Physical Machine is lightly loaded it indicates the inefficiency in resource utilization. That is the system is in under-utilized state. On the other hand, the overloaded Physical Machines leads to decrease in Quality of Service. This overloaded condition of a Physical Machine is known as HOTSPOT. In other words, Hotspot can also be defined as the state when the performance of a system falls below the minimum acceptance level. During the occurrence of Hotspot, the downtime of the Physical Machine will decrease. That is the performance of the system may go down. This may sometimes leads to interruption in user actions. The Physical Machine, to which the virtual machine has to be migrated, should be selected in such a way that upon migration the Physical Machine where the virtual machine is migrated should not be overloaded. That is after migration of the virtual machine to the Physical Machine the system’s performance level should not degrade at any time.

**4. PROPOSED METHOD:**

4.1. IQR ALGORITHM

To migrate a virtual machine from one Physical Machine to another Physical Machine, the overloaded system should be identified. To find the overloaded virtual machine, we proposed an Inter Quartile Range algorithm (IQR Algorithm). This algorithm computes the threshold value which is used to identify the virtual machine to be migrated. The flow of the algorithm begins by taking the CPU utilization of each and every virtual machines as input. These values are sorted in ascending order. The sorted array of CPU utilization values are divided into two equal halves. The middle value of the sorted array is known as Quartile2(Q2). The center element of the first half, that is from first element to Quartile2, is Quartile1(Q1). On other hand, the center element of the second half, that is from Quartile2 to last element, is Quartile3(Q3). The difference between Quartile3 and Quartile1 (Q3 - Q1) gives the Inter Quartile Range value (IQR value). The IQR value computed is multiplied with the safety parameter (s). This multiplication is done to avoid deviation in threshold. For IQR algorithm, the value of safety parameter is 1.5. This multiplied value is subtracted from 1. The resultant value gives the Threshold value. It is then compared with each CPU utilization values. The virtual machine with the CPU utilization value that is greater than the Threshold value computed is considered as the overloaded virtual machine. This overloaded virtual machine degrades the system performance. Hence the overloaded virtual machine should be migrated to an under-utilized system. While this migration the downtime of the system should be less and the performance of the system is maintained.

4.2. XEN AS HYPERVISOR

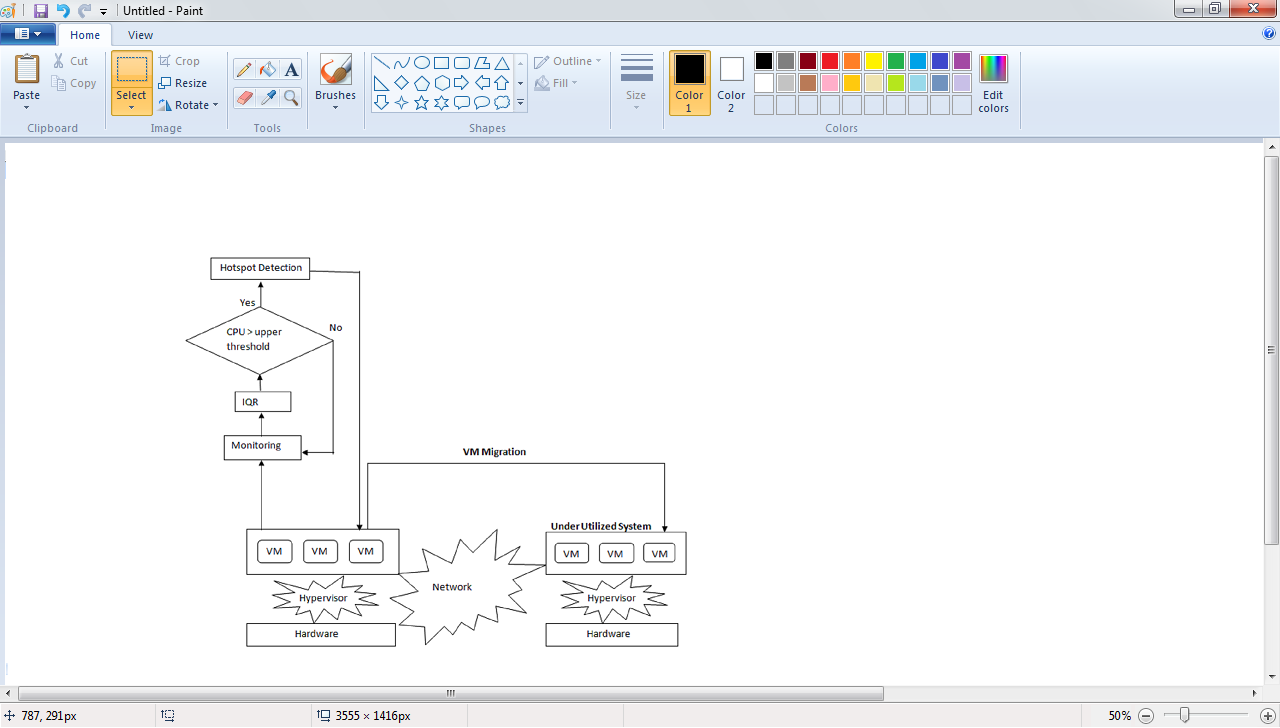
Xen creates a Virtual Machine Monitor (VMM) also known as a hypervisor. It provides a services that allow multiple computer operating systems to execute on the same computer hardware simultaneously. The hypervisor directs a hardware access and coordinating requests from the guest operating systems. Xen hypervisor supports three primary types of virtualization. They are para-virtualization, OS virtualization and full virtualization. These operating systems are aware that they are being virtualized and as such not need a virtual hardware devices, instead they make special calls to the hypervisor that allow them to access CPUs, storage and network resources. Xen runs in a more privileged CPU state than any other software on the machine. It supports for virtual machine live migration from one host to another allows workload balancing and the avoidance of down time. Users can “Live Migrate” Xen virtual machines between physical hosts across an network without loss of availability. During this process iteratively copies the memory of virtual machine to the destination without stopping its execution. Xen can be transport in a dedicated virtualization platform, such as Citrix XenServer Enterprise Edition. It Supports multiple guest operating systems such as Linux, Windows, NetBSD. It also Supports multiple Cloud platforms like CloudStack, OpenStack.XenServer provides the best-in-class performance for application and desktop virtualization with an integrated virtualization platform for Citrix [XenApp](https://www.citrix.com/products/xenapp.html) and [XenDesktop](https://www.citrix.com/products/xendesktop.html). XenServer maximizes VM density. XenIncreasing availability of Virtual Machines by using High Availability to configure policies that restart VMs on another.Consolidating multiple VMs onto physical servers reducing the number of separate disk images that need to be managed.Allowing for easy integration with existing networking and storage infrastructures.

4.3. TIME SERIES BASED ANALYSIS

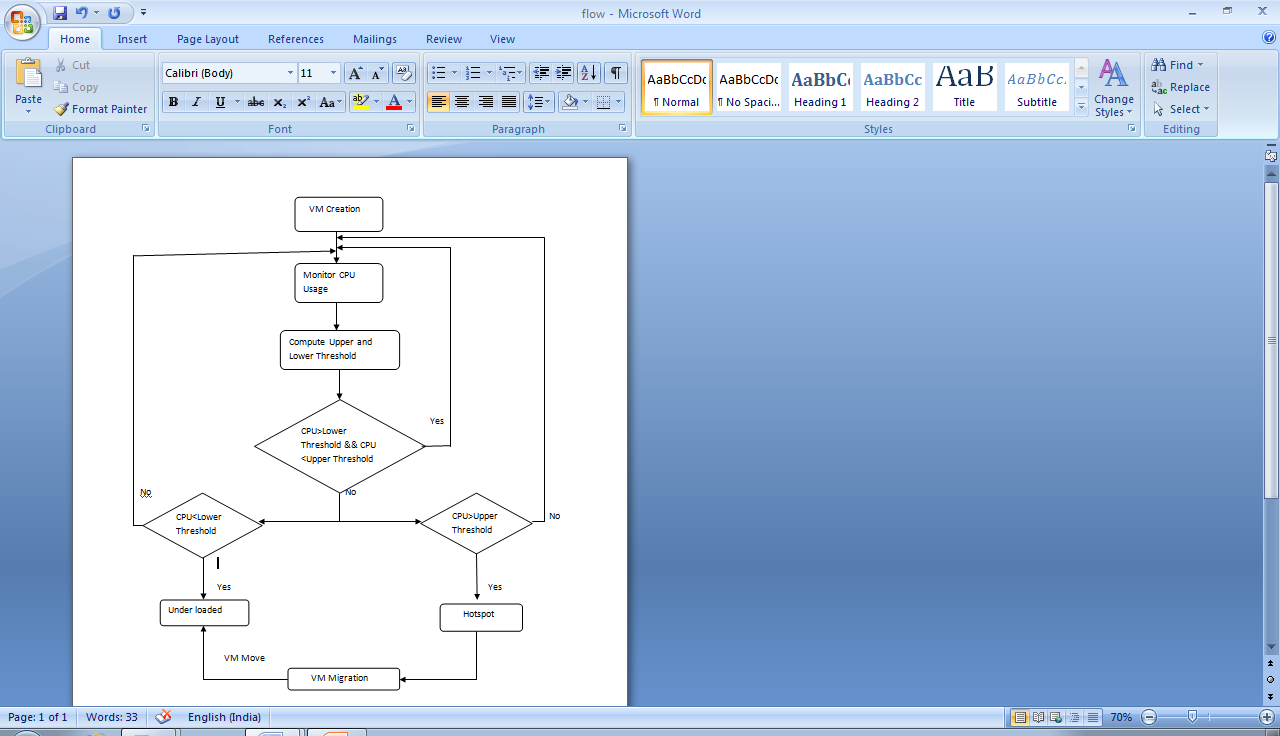
A Time Series based Analysis technique is used to periodically monitor all the Virtual Machines in a Physical Machine. This analysis is made to find the virtual machine to be migrated through the pattern observed, that is a number of virtual machines a created in a Physical Machine. The CPU usage of all the virtual machines will not be same. The Virtual machine with high CPU usage at the present state will be comparatively less in some point of time. So at a certain time period, Say 10mins, is fixed in prior. The CPU usage of all the virtual machines are monitored every 10mins. After 10mins, the Virtual Machine to be migrated is determined using the IQR algorithm. If any of the Virtual Machine has a CPU value above the threshold value, the output of IQR algorithm, that Virtual Machine will be considered as the overloaded Virtual Machine and it should be migrated.

**5. DESIGN OF THE PROJECT:**

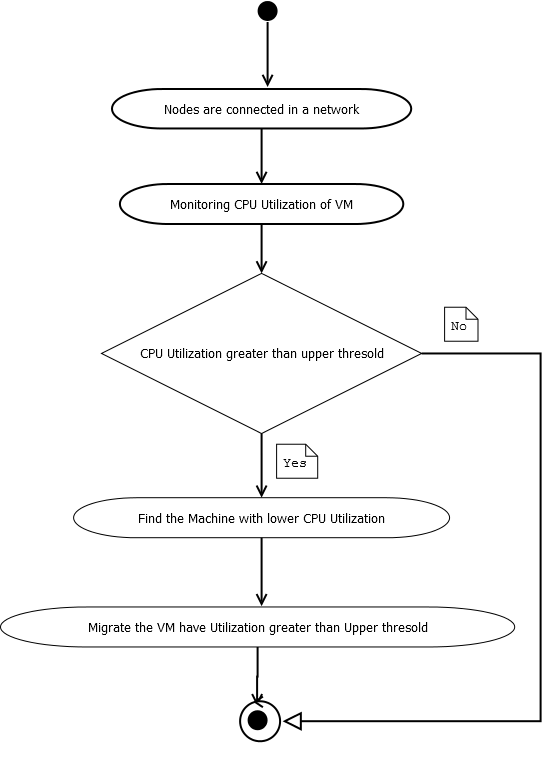
5.1. ARCHITECTURE DIAGRAM:



5.2. DATA FLOW DIAGRAM:



5.3. STATE TRANSITION DIAGRAM:



**6. SCREENSHOTS:**

**7. CONCLUSION AND FUTURE WORK:**

7.1. CONCLUSION:

In this paper, we proposed A Time Series based Hotspot detection techniques to periodically monitor the CPU usage of all the Virtual Machines in a Physical Machine. To achieve this XEN hypervisor is used. Using an Inter Quartile Range Algorithm the Threshold value is computed, which acts as a decision factor. Based on this decision factor, Server decides which Virtual Machine should be migrated. However, the downtime is high.

7.2. FUTURE WORK:

Our future work is to minimize the downtime at the time of Virtual Machine migration from one Physical Machine to another Physical Machine. In addition to CPU usage, Network bandwidth and Memory of Remote Physical Machine will be considered for Hotspot detection and Virtual Machine migration.

**8. REFERENCES:**

[1] M. Armbrustet al., “A View of Cloud Computing”, Commun. ACM, vol. 53, no. 4, 2010, pp. 50–58.

[2] Akoush, S.Sohan, R.Rice, A.Moore, A.Hopper, 2010. Predicting the performance of virtual machine migration.In:Proceedings of IEEE/ACMMASCOTS, MiamiBeach, FL, USA, doi:10.1109/MASCOTS.2010.13.

[3] V.Medina and J. M. Garcıa, “A survey of migration mechanisms of virtual machines,” ACM Computing Surveys, vol. 46, no. 3, p.30, 2014.

[4] M. Mishra, A. Das, P. Kulkarni, and A. Sahoo, “Dynamic resource management using virtual machine migrations,” IEEE Communications Magazine, vol. 50, no. 9, pp. 34–40, 2012.

[5] W. Zhao, Z.Wang, and Y. Luo, “Dynamic memory balancing for virtual machines”, ACMSIGOPS Operating Systems Review, vol.43, no. 3, pp. 37–47, 2009.

[6] A. Beloglazov and R. Buyya, “Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers,” in Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science, p. 4, 2010.

[7] A. Khan, X. Yan, S. Tao, and N. Anerousis, “Workload characterization and prediction in the cloud: a multiple time series approach,” in Proceedings of the IEEE Network Operations and Management Symposium (NOMS ’12), pp. 1287–1294, Maui, Hawaii, USA, April 2012.

[8] Y.Wu and M. Zhao, “Performancemodeling of virtual machine live migration,” in Proceedings of the IEEE 4th International Conference on Cloud Computing (CLOUD ’11), pp. 492–499, IEEE,Washington, DC, USA, July 2011.

[9] W.Hu, A. Hicks, L. Zhang et al., “Aquantitative study of virtual machine live migration,” in Proceedings of the ACM Cloud and Autonomic Computing Conference, pp. 1–11, 2013.

[10] A. B. Nagarajan, F. Mueller, C. Engelmann, and S. L.Scott.Proactive fault tolerance for HPC with Xen virtualization. In Proceedings of the 21st Annual International Conference on Supercomputing, pages 23–32, 2007.

[11] C. Jo, E. Gustafsson, J. Son, and B. Egger. Efficient live migration of virtual machines using shared storage. In Proceedings of the 9th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, pages 41–50, 2013.

[12] T. Hirofuchi, H. Nakada, S. Itoh, and S. Sekiguchi.Enabling instantaneous relocation of virtual machines with a lightweight VMM extension. In Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pages 73–83, 2010.