ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

A Survey of Energy Aware Scientific Workflows Execution Techniques in Cloud

Vishal Sharma[1], Rohit Kumar[2]

Research Scholar, Dept. of CSE, U.I.E.T., Panjab University, Chandigarh, India[1]

Assistant Professor, Dept. of CSE, U.I.E.T., Panjab University, Chandigarh, India[2]

ABSTRACT: Cloud Computing is being used widely all over the world by many IT companies as it provides various benefits to the users like cost saving and ease of use. But with the growing demands of users for computing services, cloud providers are encouraged to deploy large datacenters which consume very high amount of energy. High energy consumption increases the total cost of ownership (TCO) for service provider. Due to advantages and properties of cloud computing it is becoming a cost and performance efficient platform for executing scientific workflows. This paper reviews energy efficient approaches in cloud datacenters using virtualization and energy aware execution of scientific workflows.

KEYWORDS: Energy efficiency, scientific workflows, VM consolidation, VM migration,

I. INTRODUCTION

Cloud computing now a day is paradigm shifting technology. Cloud computing platforms are used as the infrastructure for flexibility allowing business to scale according to user demands [1]. It changes the way an IT company operates these days. Cloud computing is consumer-service delivery model in which IT services, which includes storage space, computing or processing power, network are provided to the users/organizations through internet. The services are billed based on the amount of their usage and users/organizations are charged accordingly to the pre decided norms and conditions which are settled between the two parties known as SLA (Service Level Agreement). Cloud computing gives organizations the freedom to concentrate on their business goals and customers rather on hardware technology or computing devices. A start up with a innovative idea can now easily experiment its business by lending resources from cloud service providers and if their business expands they can easily scale up the resources needed. Cloud computing has many inherent qualities and all we can use the services using a internet connection from anytime anywhere. Cloud computing leverages IT services to the users as scalable and metered service. Virtualization in the cloud datacenters provide opens up enormous way to save energy by consolidating virtual machines efficiently in order to use resources optimally. A datacenter consists of many components and energy can be saved by applying techniques and policies at each component level. CPU and memory are recognized as main components that consume most amount of energy. Cooling is also energy consuming and different techniques by which it can be addressed are using local or green source of electricity, establishment of datacenters in cold climate, thermal management etc. CPU energy consumption is reduced by using virtualization and consolidating large amount workload on single physical machine. But this can give performance degradation and SLA violations if not used properly. VM migration is used to handle VMs optimally. Using VM migration underloaded host can be switched off or put to sleep mode after migrating VMs running on that host and hence, saves energy. VM migration is of 2 types Precopy migration and Postcopy migration. Research work showed that Postcopy VM migration technique is efficient and thus is used over Precopy VM migration. VM migration itself has some energy overheads hence a migration condition is used by researchers to reduce number of migrations.
Workflows is defined as many coarse-grained parallel tasks that are processed or executed in a specific order to accomplish a real application. Many scientific analysis can be expressed easily as a workflow. In Scientific workflows, heavy volume of data processing is required and workflows with few million tasks are not unusual.
Cloud computing gives a new approach on how to deploy and execute different scientific workflows in cloud and becomes a platform for cost and time effective execution.

A workflow is conventionally demonstrated as a Directed Acyclic Graph (DAG).

- Node in the DAG signifies a workflow task

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

- Edges signify dependency between the tasks that pressure the order in which the tasks are performed.

A simple DAG with four tasks is shown in figure 1 below.

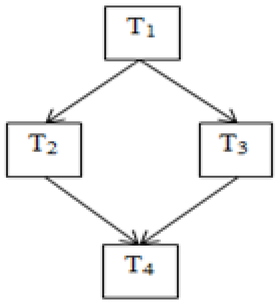

Figure 1: Directed Acyclic Graph

CloudSim is used to simulate the cloud environment and to test the proposed approaches. For workflow scheduling an extension of CloudSim is used named WorkflowSim. WorkflowSim is an open source toolkit which simulates workflow execution in virtualizes cloud datacenter environment. All energy efficient techniques like DVFS, virtual machine migration are used to save energy while maintaining the deadline constraint of workflows. Different real time workflows used by researchers and each has a particular structure. Figure 2 shows the structures of four scientific workflows Montage, LIGO, SPIHT and CyberShake.

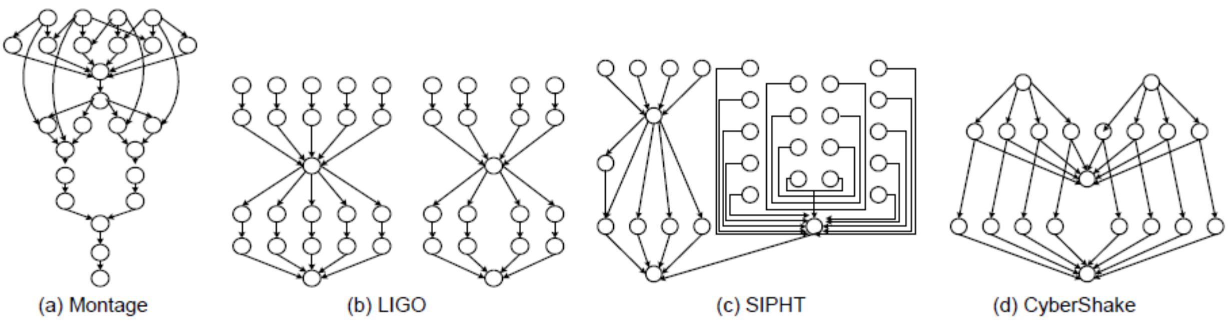

Figure 2: Structure of workflows

II. RELATED WORK

In this section we will discuss the main energy efficient algorithms and policies presented by different researchers firstly in virtualized cloud and then specially for energy efficient workflow scheduling in cloud datacenters.

In [2], authors address the problem energy efficient resource management in large sacale virtualized data centers. Nathuji et.al. are to first to study power management in context of virtualized datacenters. In 1 authors presents an approach called VirtualPower. In his approach both hardware power scaling and software based methods are used to control the power consumption of the underlying platform. They also consolidate the virtual machines using migration to achieve energy efficiency. Both Soft and hard actions are combined to decrease power utilization. Soft actions are CPU idling and throttling and hard actions are DVFS and consolidation. Authors split resource management into local and global policies. At local level system uses the power management policies of the guest VMs at physical node. At

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

global level energy efficient policies are implemented at physical machines and at hardware infrastructure. Results of their approach is very promising and they reduce the power consumption by 34% by using VirtualPower.

Song et. al. studied the problem of energy efficiency in multi application virtualized datacenters. In [3], authors proposed a multi tier resource scheduling scheme that provides on demand services to clients requests through the resources acquired among VMs. When competition among resources occur, critical services are given preferences over other low priority services. Authors proposed a prototype RAINBOW which works at three levels. First is the application level scheduler second is local level scheduler and third is global level scheduler. During resource utilization decisions CPU and RAM utilizations are taken into considerations. Experimental results shows that using RAINBOW 9-16% improvements was achieved in performance of critical services.

In [4] authors work was focused on the effort to coordinate all the solutions that are available at hardware and software level and policies at local and global level to achieve required energy efficiency at datacenters. Authors proposed a system where they combine five different power management techniques and coordinate among those. Authors formulate this problem as control theory coordinate among these using feedback control loops. Five controllers are used which are (i) efficiency controller (ii) server manager (iii) enclosure manager (iv) group manager (v) VM controller. Efficiency controller optimizes power consumption for individual server by adjusting P states of CPU. Server manager applies power strategies at server level. Enclosure manager and group manager Improves power efficiency at enclosure and datacenter level respectively. Authors simulated their coordinated approach and achieved 64% power reduction when compared to non coordinated approaches.

In [5], authors investigate the problem of dynamic placement of application in virtualized systems, while minimizing the energy losses and SLA violations. Authors also take migration costs and power utilization into account for calculating overall results. Authors presented a pMapper architecture uses placement algorithms to solve the problem of minimizing power consumption under fixed performance requirements. pMapper consists of three managers and an arbitrator which coordinates the actions of these managers. These are performance manager, power manager and migration manager. Power manager adjusts hardware power states and apply DVFS. Performance manager monitors the workload and resize VMs accordingly. Migration manager handles and issues instructions for live migration of virtual machines. This method is a continuous optimization approach in which at each time frame VM placements were optimized to minimize power consumption and maximized performance.

The algorithm proposed in [6], focused on minimizing energy consumption by scheduling virtual machines in the system efficiently. The main parameters that had been focused were DVFS and active cooling. Another features like shutdown of underutilized machines and migrations of workloads from the machines that are operating below a specific threshold were also used. Authors give each node an energy efficiency priority based on the ratio of mips and energy utilized by it. The algorithm was capable of performing the scheduling of VMs in non federated, homogeneous and heterogeneous datacenters. The algorithm was able to improve power consumption in clouds significantly when used in heterogeneous datacenters.

In [7], anton et.al. investigates the problem of high energy consumption in large virtualized data centers. Authors use virtual machine migration and switching idle nodes to sleep mode to conserve energy. Their objective was to give QoS the users and reducing the total cost of ownership for the providers. They tried to reduce the number of migrations due to migration overheads. Authors presented online algorithm for energy and performance efficient VM placement in cloud data center. Algorithms are presented which were heuristic based are presented. First algorithm determines which host is overloaded and which host is underloaded. They used thresholding technique for this. Overloaded hosts causes performance degradation hence, VMs are migrated from overloaded hosts. Underloaded hosts are switched to sleep mode after migrating their VMs. Second algorithm finds a energy efficient and suitable placement for these VMs. The algorithms can work for any type of workload. simulation is done using cloudsim simulator and experimental results showed that they reduced the energy and carbon dioxide emission and also follow the strict SLA.

The approach presented in [8] also focused on VM scheduling in order to get an energy efficient system. Authors had proposed a combination of two algorithms: allocation algorithm used for allocating the jobs and a migration algorithm for optimal migration of VMs considering minimum migrations and minimum energy consumption. The objective of allocation algorithm is to reduce the number of servers used to execute the workload and increasing the number of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

servers which can be put in sleep mode. It used bin packing problem approach and it was compared with best fit algorithm. The authors had proposed that by using this approach of combining the allocation algorithm with migration algorithm in a linear integer program, a significant amount of energy can be saved depending upon system loads. Experimental results show that by using exact allocation and migration algorithm 10-50% servers were idle depending on the workload that can put to sleep mode. At low loads results were quite significant.

In [9], authors worked on the problem of energy efficient virtual machine management in virtualized cloud data centers. authors worked on the VM allocation and reallocation in cloud to efficiently utilize the resources while reducing the number of migrations and overhead incurred by them. The objective of their work was to reduce energy consumption, reducing the migration number, maintaining the QoS and to complete their solution in computational time. Authors used OpenNebula platform and studied the migration behaviour on that platform and implemented host management strategy to effectively consolidate the VMs. Authors proposed a VM management approach named SOPVP. Two lists are maintained for each host one for the hosts that have more CPU than memory and other for the hosts that have more memory than CPU. VMs are allocated to the applications based on their CPU or memory requirements. Proposed algorithm SOPVP was compared with other default algorithm provided in OpenNebula and results showed that their algorithm saved significant energy while maintaining the QoS.

Mohammad et. al. in [10] proposed an energy aware virtual machine migration algorithm which migrates the virtual machine on the criteria of load on the servers, number of VMs, hit count and other important factors. Firstly a victim server underutilized server is selected based on the above factor and then target server is selected to place the VMs in such a way so as to reduce migration oscillation. They also used a global threshold for datacenter after which more servers were made active to accumulate workload. They proposed EAM algorithm and simulated it on the presented below:

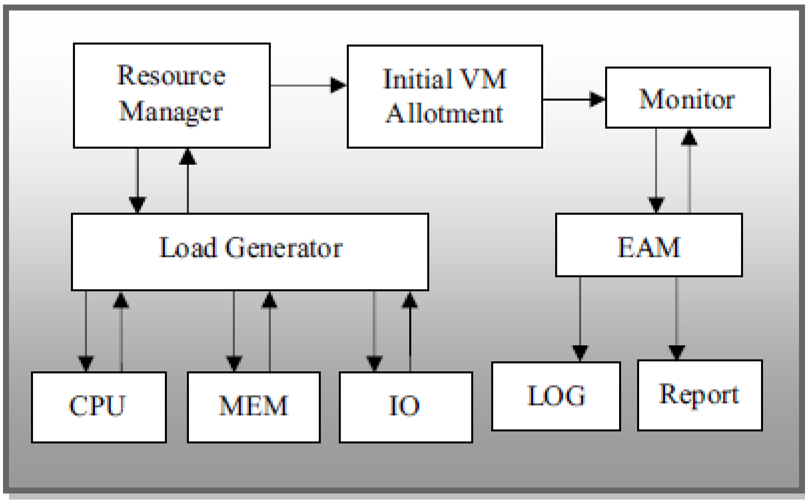

Figure 3: EAM algorithm

Continuous monitoring was used to make good decisions on migration of VMs. Experimental results showed that about 20% energy is reduced using EAM in both static and dynamic load scenarios.

In [11], authors presented a energy aware consolidation framework pSciMapper for scientific workflows. They presented consolidation as hierarchical clustering problem. Kernel Canonical Correlation Analysis (KCCA) is used to map the resource requirement to performance and energy consumption. pSciMapper is implemented on two real scientific workflows and three synthetic workflows. Experimental results show that authors were able to reduce the total power consumption by 56%. Their consolidation algorithm incurs low overheads and is suitable for large scale workflows.

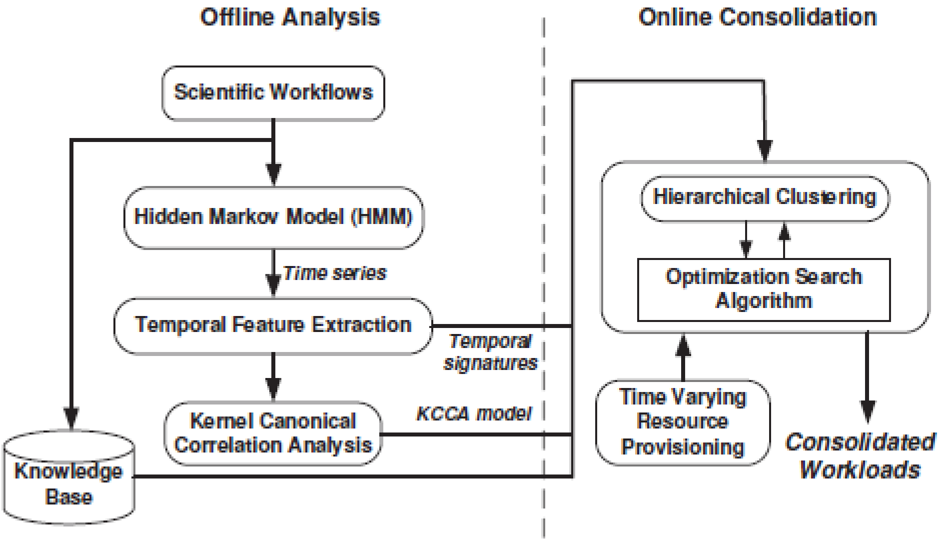

Figure 4: pSciMapper scheme

In [12], authors proposed two algorithms for scheduling of scientific workflows in cloud environment to achieve time and energy efficiency. First algorithm they proposed is EHEFT (Enhancing Heterogeneous Earliest Finish Time) which finds the set of processors that were executing the tasks inefficiently and shut them down to reduce energy consumption. EHEFT had three phases task scheduling phase, discovering inefficient processors phase and rescheduling phase. Second algorithm proposed was ECPOP (Enhancing Critical Path On Processor) which reduce the power consumption by increasing the utilization of task execution while maintaining the makespan. It also had three phases task scheduling phase, discovering non-critical processor phase and rescheduling phase. Experimental results showed that enhanced algorithms gave better results than the base algorithms.

In [13] authors presented an energy efficient task scheduling algorithm for scientific workflows using frequency scaling and strictly following user deadline. Their work focused on heterogeneous machines with different task runtime and frequency capabilities. Initially tasks of the workflows re mapped to available machines. Energy Aware Stepwise Frequency Scaling (ESFS) algorithm determines the processor frequency to be used for the execution of each task taking into account its deadline and ensure that deadline was not violated. The proposed algorithm works iteratively to gradually scale the processor frequency used for each task for as long as overall energy savings are increased. The next available frequency mode for each processor is used as a lower bound in each iteration to scale the frequency of the tasks. ESFS algorithm is implemented on LIGO,SPIHT and Montage workflows and results showed that energy consumed is reduced when compared with baseline algorithms HEFT and EES.

In [14] authors studied the studied the problem of energy consumption and carbon dioxide emission at data center level while executing scientific workflows. They proposed an algorithm which uses DVFS and limiting the number of VM migrations to reduce energy consumption and high resource utilization rate. They used five step workflow scheduling algorithm. The five steps they used are in step 1 earliest completion time estimation phase . Step 2 is the DVFS phase . Step 3 is data center selection phase which selects the data center from the list of data centers selected in first step. Step 4 is the forward workflow scheduling phase and step 5 is the backward workflow scheduling phase. Experimental results showed that they achieve upto 30% energy saving which gives high profit and less CO_2 emission.

In [15] authors presented an energy aware allocation method EnReal for the execution of scientific workflows in cloud datacenters. Authors firstly presented an energy model to measure energy consumption by workflows in cloud platform. Total energy consumed is given by sum of energy consumption for application execution and dynamic executions. Secondly a four step algorithm EnReal is presented. In step 1 the task requests of scientific workflows are portioned by requested start time. In step 2 resource monitoring of physical machines was done. In step 3 migration of VMs was done from idle PMs. in step 4 an energy-aware global resource allocation policy is designed to deploy VMs dynamically for scientific workflow executions. Simulation was done using CloudSim toolkit and results showed that algorithm gives effective results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

In [16] authors addressed the problem of energy consumption during workflow execution in cloud environment. They proposed a multiple procedure scheduling algorithm with objectives to maximize the resource utilization, reducing energy consumption and guarantee the deadlines of scientific workflows. Algorithm designed by authors focused on the three factors that consume energy in cloud i.e. VM power to execute job modules, the overhead of instantiating and tearing down VMs and communication cost in task migration. Their heuristic scientific workflow scheduling algorithm consists of three procedures. First procedure was eTimeBound which estimates the execution time of running a workflow in selected data centers. Second procedure was ForwadMapping which selects the highest priority task from the workflow based on critical path of the workflow. Third procedure was backward mapping which enforces the execution in such a way so that energy is minimized and workflow tasks also execute under their deadlines. CPU usage was restricted between two thresholds to minimize the energy consumption. Authors simulated a datacenter with 100 servers and compared their proposed algorithm with three existing algorithms. Results obtained showed that they achieved better resource utilization and lowest energy consumption.

III. CONCLUSION AND FUTURE WORK

In this paper, we have conducted the survey on algorithms and policies for reducing energy consumption in virtualized cloud datacenters. Starting from the first work by nathuji et.al. Virtual machine consolidation and migration is the main technique to reduce energy consumption. At node level DVFS is also used by most of the researchers as it uses dynamic scaling for energy consumption reduction. Presently cloud platform is best platform for executing scientific workflows due to its inherent characteristics like auto scaling, on demand access and resource pooling. Workflows contains large amount of parallel tasks and can consume huge amount of time and energy if they are not scheduled efficiently. Different techniques are presented here and discussed in brief.

In the future, we will propose an energy efficient algorithm for real scientific workflow scheduling in cloud. That will present a VM migration criteria based on energy consumption of each VM and number of tasks a VM is executing presently.

REFERENCES

1. Borgetto, Damien, and Patricia Stolf. "An energy efficient approach to virtual machines management in cloud computing." Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on. IEEE, 2014.
2. Nathuji, Ripal, and Karsten Schwan. "VirtualPower: coordinated power management in virtualized enterprise systems." In ACM SIGOPS Operating Systems Review, vol. 41, no. 6, pp. 265-278. ACM, 2007.
3. Song, Ying, Hui Wang, Yaqiong Li, Binquan Feng, and Yuzhong Sun. "Multi-tiered on-demand resource scheduling for VM-based data center." InProceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, pp. 148-155. IEEE Computer Society, 2009.
4. Raghavendra, Ramya, Parthasarathy Ranganathan, Vanish Talwar, Zhikui Wang, and Xiaoyun Zhu. "No power struggles: Coordinated multi-level power management for the data center." In ACM SIGARCH Computer Architecture News, vol. 36, no. 1, pp. 48-59. ACM, 2008.
5. Verma, Akshat, Puneet Ahuja, and Anindya Neogi. "pMapper: power and migration cost aware application placement in virtualized systems." InMiddleware 2008, pp. 243-264. Springer Berlin Heidelberg, 2008.
6. Lago, Daniel Guimaraes do, Edmundo RM Madeira, and Luiz Fernando Bittencourt. "Power-aware virtual machine scheduling on clouds using active cooling control and DVFS." In Proceedings of the 9th International Workshop on Middleware for Grids, Clouds and e-Science, p. 2. ACM, 2011.
7. Beloglazov, Anton, and Rajkumar Buyya. "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers." Concurrency and Computation: Practice and Experience 24, no. 13 (2012): 1397-1420.
8. Ghribi, Chaima, Makhlouf Hadji, and Djamal Zeghlache. "Energy efficient vm scheduling for cloud data centers: Exact allocation and migration algorithms." In Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on, pp. 671-678. IEEE, 2013.
9. Borgetto, Damien, and Patricia Stolf. "An energy efficient approach to virtual machines management in cloud computing." Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on. IEEE, 2014.
10. Al Shayeji, Mohammad H., and M. D. Samrajesh. "An energy-aware virtual machine migration algorithm." In Advances in Computing and Communications (ICACC), 2012 International Conference on, pp. 242-246. IEEE, 2012.
11. Zhu, Qian, Jiedan Zhu, and Gagan Agrawal. "Power-aware consolidation of scientific workflows in virtualized environments." In Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1-12. IEEE Computer Society, 2010.
12. Thanavanich, Thanawut, and Putchong Uthayopas. "Efficient energy aware task scheduling for parallel workflow tasks on hybrids cloud environment." InComputer Science and Engineering Conference (ICSEC), 2013 International, pp. 37-42. IEEE, 2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

13. Pietri, Ilia, and Rizos Sakellariou. "Energy-aware workflow scheduling using frequency scaling." In Parallel Processing Workshops (ICCPW), 2014 43rd International Conference on, pp. 104-113. IEEE, 2014.
14. Cao, Fei, and Michelle M. Zhu. "Energy Efficient Workflow Job Scheduling for Green Cloud." In Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International, pp. 2218-2221. IEEE, 2013.
15. Xu, Xiaolong, Wanchun Dou, Xuyun Zhang, and Jinjun Chen. "EnReal: An Energy-Aware Resource Allocation Method for Scientific Workflow Executions in Cloud Environment."
16. Khaleel, Mustafa, and Michelle M. Zhu. "Energy-Aware Job Management Approaches for Workflow in Cloud." In Cluster Computing (CLUSTER), 2015 IEEE International Conference on, pp. 506-507. IEEE, 2015.