EFFICIENT COST AND ENERGY MANAGEMENT IN CLOUD DATA CENTERS

Priyanka uttarapally[1],Deepika sheri[2],Belli sowmya sri[3]

[1]Associate Professor,M.Tech.,BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (v), hayath nagar (m), r.r dt. Hyderabad

[2]Associate Professor,M.Tech.,BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (V), Hayath Nagar (M), R.R Dt. Hyderabad

Department of CSE (DS),

[3]UG Students, BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (V), Hayath Nagar (M), R.R Dt. Hyderabad

ABSTRACT

We target the problem of managing the power states of the servers in a Cloud Data Center (CDC) to jointly minimize the electricity consumption and the maintenance costs derived from the variation of power (and consequently of temperature) on the servers' CPU. More in detail, we consider a set of virtual machines (VMs) and their requirements in terms of CPU and memory across a set of Time Slot (TSs). We then model the consumed electricity by taking into account the VMs processing costs on the servers, the costs for transferring data between the VMs, and the costs for migrating the VMs across the servers. In addition, we employ a material-based fatigue model to compute the maintenance costs needed to repair the CPU, as a consequence of the variation over time of the server power states. After detailing the problem formulation, we design an original algorithm, called Maintenance and Electricity Costs Data Center (MECDC), to solve it. Our results, obtained over several scenarios from a real CDC, show that MECDC largely outperforms two reference algorithms, which instead either target the load balancing or the energy consumption of the servers.

I. INTRODUCTION

Embarking on the quest to redefine Cloud Data Centre management, our Java-based project stands at the forefront of innovation. Our primary objective revolves around harmonizing maintenance expenditures and electricity consumption. Leveraging the versatility of Java, we're engineering a comprehensive solution that meticulously balances these crucial factors. Through sophisticated algorithms and smart resource allocation strategies, our project aims to redefine operational paradigms, ensuring optimal resource utilization while curbing unnecessary costs. By fusing cutting-edge technology with sustainable practices, we're driving towards a future where Cloud Data Centres operate efficiently, economically, and with a reduced environmental footprint.

At the core of our project lies a dedication to revolutionize data infrastructure management. Our Java-centric approach isn't just about coding—it's about pioneering a paradigm shift. By integrating novel methodologies, we're creating a robust framework that empowers Cloud Data Centres to thrive economically without compromising performance. Join us in shaping a landscape where cost-efficiency and environmental consciousness converge seamlessly, redefining the benchmarks of Cloud Data Centre operations.

II. LITERATURE SURVEY

Cloud computing has promoted the success of big data applications such as medical data analyses. With the abundant resources provisioned by cloud platforms, the QoS (quality of service) of services that process big data could be boosted significantly. However, due to unstable network or fake advertisement, the QoS published by service providers is not always trusted. Therefore, it becomes a necessity to evaluate the service quality in a trustable way, based on the services' historical QoS records. However, the evaluation efficiency would be low and cannot meet users' quick response requirement, if all the records of a service are recruited for quality evaluation.

Moreover, it may lead to 'Lagging Effect' or low evaluation accuracy, if all the records are treated equally, as the invocation contexts of different records are not exactly the same. In view of these challenges, a novel approach named Partial-HR (Partial Index Terms—big data, cloud, context-aware service evaluation, historical QoS record, weight Historical Records-based service evaluation approach) is put forward in this paper. In Partial-HR, each historical QoS record is weighted based on its service invocation context. Afterwards, only partial important records are employed for quality evaluation. Finally, a group of experiments are deployed to validate the feasibility of our proposal, in terms of evaluation accuracy and efficiency.

The existing work either only considers partial context elements, or lacks quantitative weight model for historical QoS records. Therefore, it becomes a challenging task to develop a quantitative weight model that considers all the context elements, for evaluating the quality of big data services accurately and efficiently. In view of this challenge, a novel service evaluation approach Partial-HR is proposed in this paper. Partial-HR not only considers all the important context elements of service invocation (i.e., invocation time, input size and user location), but also satisfies the Volatility Effect and Marginal Utility. Through Partial-HR, we can select partial important historical QoS records for service evaluation, so that the evaluation accuracy and efficiency could be improved. Through a set of experiments, we validate the feasibility of our proposal.

In cloud environment, the advertised QoS information of big data services is not always trusted. Therefore, it becomes a necessity to evaluate the service quality based on historical QoS records. Today, many researchers have studied this problem and given their proposals. In the problem of QoS credibility is firstly put forward, and the historical QoS records are suggested to be considered for evaluating the real quality of service. In the literature the service's QoS credibility is calculated, by comparing the historical QoS data with the SLA (Service Level Agreement) promised by service providers. Afterwards, it became popular to utilize the historical QoS records of services for various trustable service-oriented applications, such as service recommendation, service evaluation, service selection and service composition. However, in the above literatures, the weight problem of different historical QoS records is discussed.

Due to the unstable network or fake advertisement, the QoS information of services that process big data in cloud, is not always trustable as advertised by service providers. Therefore, it becomes a necessity to evaluate the service quality in a trustable way, based on the historical QoS records. However, it may lead to low efficiency if all the records are considered in service quality evaluation. Moreover, evaluation accuracy would be low if all the historical QoS records are treated equally, as their service invocation contexts are not exactly the same. In view of these challenges, a novel evaluation approach named Partial-HR is proposed in this paper, which not only considers the service invocation context, but also satisfies 'Volatility Effect' and

‘Marginal Utility’ simultaneously. Through a set of experiments, we validate the feasibility of Partial-HR in terms of evaluation accuracy and efficiency. In the future, we will introduce more context elements into our weight model for historical QoS records, so as to further improve the evaluation accuracy of big data ser-vices in cloud.

III. EXISTING SYSTEM

Focusing on the memory and storage management,” Unified performance and power modeling of scientific workloads” employ power performance information to estimate the desired storage and memory parameters in order to preserve energy and costs in the CDC. It is important to note that their quasi-analytical performance modeling can be accurate, but it requires a deep understanding of each individual application running on the VM and the server. Therefore, a consistent amount of preliminary information is needed and, as a consequence, the pre-processing time of the problem may sensibly increase.

IV. PROPOSED SYSTEM

This context poses several challenges: What is the impact of the maintenance costs on the total costs? Is it beneficial to leverage the tradeoff between electricity consumption and maintenance costs? How to optimally formulate the problem? How to design an efficient algorithm to tackle it? The goal of this paper is to shed light on these issues. More in detail, we first present a simple (yet effective) model to compute the maintenance costs, given the variation over time of the power states for a set of servers. In addition, we adopt a detailed model to compute the power consumed by the CDC. Specifically, our power model takes into account the CPU-related electricity costs of the servers, the costs for transferring data among the servers, and the costs for migrating the Virtual Machines (VMs) running on the servers. After formulating the problem of jointly reducing the CDC electricity consumption and the related maintenance costs, we propose a new algorithm, called Maintenance Energy Costs Data Center (MECDC), to tackle it.

Future Work

As next steps, we plan to face different issues, including: i) the definition and evaluation of more complex failure models to take into account the impact on different components, as well as different temperatures of CPU cores, ii) the introduction of delay costs for migrating VMs across PSs, iii) the application of our approach to a set of CDCs, each of them subject to different electricity prices (e.g., due to different CDC locations).

V. MODULE IMPLEMENTATION

1. **Network Manager**
2. **Allocation Manager**
3. **Maintenance Cost**
4. **Electricity Cost**

Network Manager

Network Manager is a program for providing detection and configuration for systems to automatically connect to network. Network Manager's functionality can be useful for both wireless and wired networks. For wireless networks, Network Manager prefers known wireless networks and has the ability to switch to the most reliable network. Network Manager-aware applications can switch from online and offline mode. Network Manager also prefers wired connections over wireless ones, has support for modem connections and certain types of VPN.

Allocation Manager

An allocation manager (mid left part of the figure) distributes the VMs over the PSs, by ensuring that each VM receives the required amount of CPU and memory from the PS hypervisor. Focusing on the tasks performed by the allocation manager, this element is in charge of running the proposed VMs' allocation algorithm, which is able to leverage the tradeoff between electricity costs and maintenance costs by acting on the PSs power states.

Maintenance Cost

We first present a simple (yet effective) model to compute the maintenance costs, given the variation over time of the power states for a set of servers. In addition, we adopt a detailed model to compute the power consumed by the CDC. Specifically, our power model takes into account the CPU-related electricity costs of the servers, the costs for transferring data among the servers, and the costs for migrating the Virtual Machines (VMs) running on the servers. After formulating the problem of jointly reducing the CDC electricity consumption and the related maintenance costs, we propose a new algorithm, called Maintenance Energy Costs Data Center
(MECDC) to tackle it.

Electricity Cost

The electricity costs as the sum of three different contributions: i) the data processing costs on the PSs, ii) the data transferring costs among the VMs located on different PSs, and iii) the costs for migrating the VMs across different PSs. The following subsections detail the different cost components. According to which the power consumption of each PS in AM is proportional to the CPU utilization due to data processing tasks running on the hosted VMs. On the other hand, when the PS is in SM, we assume that its power consumption is negligible. We then consider the electricity costs derived from the exchange of data between VMs running on different PSs. we assume that the total costs due to data transferring are the sum of a static term, which considers the power consumed by the network interfaces of the PS, and a linear one, which instead takes into account the amount of data transferred between VMs.

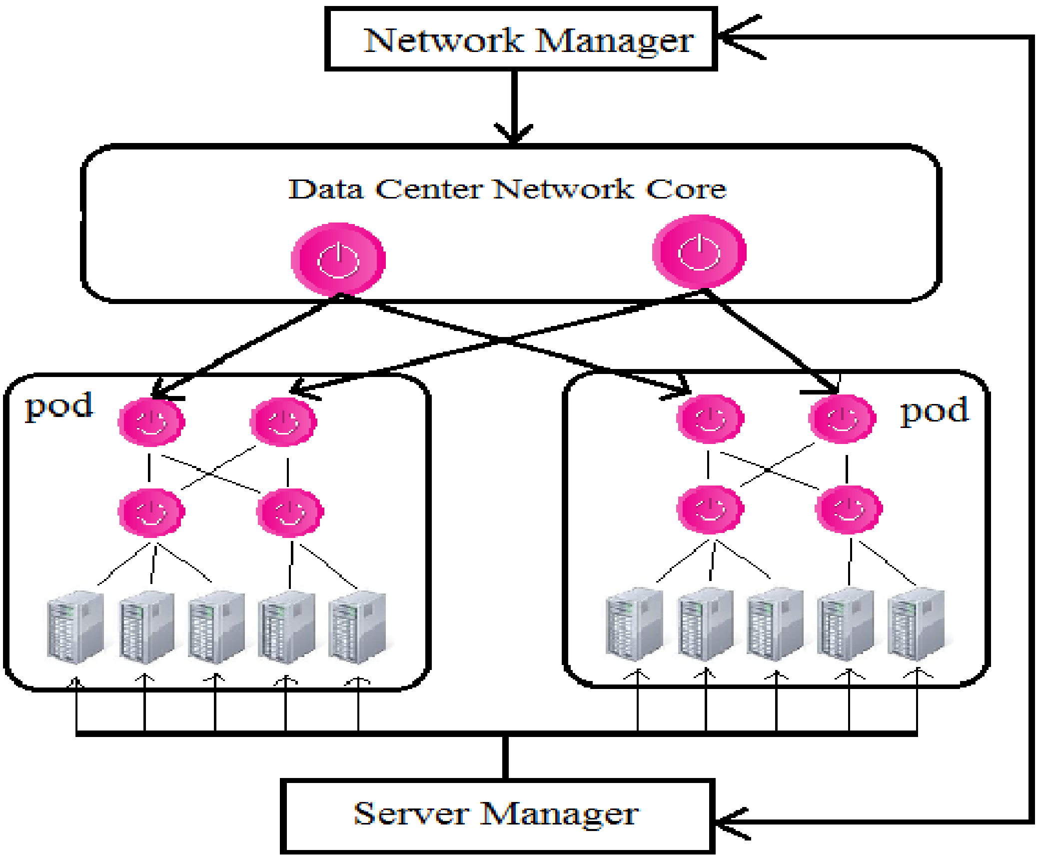

Fig1: Architecture

VI. ALGORITHM

1.Maintenance Energy Costs Data Center (MECDC)

Specifically, our power model takes into account the CPU-related electricity costs of the servers, the costs for transferring data among the servers, and the costs for migrating the Virtual Machines (VMs) running on the servers. After formulating the problem of jointly reducing the CDC electricity consumption and the related maintenance costs, we propose a new algorithm, called Maintenance Energy Costs Data Center (MECDC), to tackle it.

VII. CONCLUSION

We have targeted the problem of jointly managing the maintenance costs and the electricity consumption in a CDC. After showing that changing the power states of PSs has an impact on both the failure management costs, as well as the energy consumption, we have formulated the OMEC problem, with the goal of jointly managing the aforementioned costs. Since the OMEC problem is NP Hard, we have described the MECDC algorithm, which has been designed to wisely leverage the tradeoff between different costs, as well as taking into account their long term impact over time. Results, obtained over a set of realistic scenarios, clearly show that MECDC always requires consistently lower costs compared to the FFD and NFD reference algorithms. Moreover, we have also shown that the total costs obtained by MECDC are also close to a lower bound. In addition, the computation time, obtained from a scenario in which there are hundreds of VMs and by running the algorithm on a Desktop PC, is very low, i.e., less than 2 [s] on average.

VIII. REFERENCES:

1. H. Wells, World Brain, London, United Kingdom:Methuen & Co., 1938.
2. V. Bush, et, "As we may think", *The Atlantic Monthly*, vol. 176, no. 1, pp. 101-108, 1945.
3. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, et al., "A view of cloud computing", *Commun. ACM*, vol. 53, no. 4, pp. 50-58, 2010.
4. P. Mell and T. Grance, et, "The NIST definition of cloud computing", June 2018.
5. M. Hilbert and P. López, "The worlds technological capacity to store communicate and compute information", *Sci.*, vol. 332, no. 6025, pp. 60-65, 2011.
6. C. D. Patel and A. J. Shah, "Cost model for planning development and operation of a data center", 2005.
7. K. Bilal, S. U. R. Malik, O. Khalid, A. Hameed, E. Alvarez, V. Wijaysekara,

et al., "A taxonomy and survey on green data center networks", *Future Generation Comput. Syst.*, vol. 36, pp. 189-208, 2014.

8. S. Srikantaiah, A. Kansal and F. Zhao, "Energy aware consolidation for cloud computing", *Proc. Conf. Power Aware Comput. Syst.*, vol. 10, pp. 1-5, 2008.
9. A. Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing", *Future Generation Comput. Syst.*, vol. 28, no. 5, pp. 755-768, 2012.
10. C. Mastroianni, M. Meo and G. Papuzzo, "Probabilistic consolidation of virtual machines in self-organizing cloud data centers", *IEEE Trans. Cloud Comput.*, vol. 1, no. 2, pp. 215-228, Jul. 2013.
11. W. Fang, X. Liang, S. Li, L. Chiaraviglio and N. Xiong, "VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers", *Comput. Netw.*, vol. 57, no. 1, pp. 179-196, 2013.
12. L. Chiaraviglio, P. Wiatr, P. Monti, J. Chen, J. Lorincz, F. Idzikowski, et al., "Is green networking beneficial in terms of device lifetime?", *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 232-240, May 2015.
13. L. Chiaraviglio, N. Blefari-Melazzi, C. Canali, F. Cuomo, R. Lancellotti and M. Shojafar, "A measurement-based analysis of temperature variations introduced by power management on commodity hardware", *Proc. Int. Conf. Transparent Opt. Netw.*, pp. 1-4, 2017.
14. D. Frear, D. Grivas and J. Morris, "Thermal fatigue in solder joints", *JOM*, vol. 40, no. 6, pp. 18-22, 1988.
15. W. Lee, L. Nguyen and G. S. Selvaduray, "Solder joint fatigue models: review and applicability to chip scale packages", *Microelectron. Rel.*, vol. 40, no. 2, pp. 231-244, 2000.
16. R. P. G. Muller, An Experimental and Analytical Investigation on the Fatigue Behaviour of Fuselage Riveted Lap Joints. The Significance of the Rivet Squeeze Force and a Comparison of 2024-T3 and Glare 3, The Netherlands:Delft Univ. Technology, 1995.
17. J. Mi, Y.-F. Li, Y.-J. Yang, W. Peng and H.-Z. Huang, "Thermal cycling life prediction of Sn-3.0 Ag-0.5 Cu solder joint using type-I censored data", *Sci. World J.*, vol. 2014, 2014.
18. M. Dayarathna, Y. Wen and R. Fan, "Data center energy consumption modeling: A survey", *IEEE Commun. Surveys Tutorials*, vol. 18, no. 1, pp. 732-794, Jan.-Mar. 2016.

19.W. Voorsluys, J. Broberg, S. Venugopal and R. Buyya, "Cost of virtual machine live migration in clouds: A performance evaluation", *Proc. IEEE Int. Conf. Cloud Comput.*, pp. 254-265, 2009.

20.J. Stoess, C. Lang and F. Bellosa, "Energy management for hypervisor-based virtual machines", *Proc. USENIX Annu. Tech. Conf.*, pp. 1-14, 2007.

21.H. Liu, H. Jin, C.-Z. Xu and X. Liao, "Performance and energy modeling for live migration of virtual machines", *Cluster Comput.*, vol. 16, no. 2, pp. 249-264, 2013.

22.G. Soni and M. Kalra, "A novel approach for load balancing in cloud data center", *Proc. IEEE Int. Advance Comput. Conf.*, pp. 807-812, 2014.

23.A. Beloglazov and R. Buyya, "Energy efficient resource management in virtualized cloud data centers", *Proc. 10th IEEE/ACM Int. Conf. Cluster Cloud Grid Comput.*, pp. 826-831, 2010.

24.J. Bi, Z. Zhu, R. Tian and Q. Wang, "Dynamic provisioning modeling for virtualized multi-tier applications in cloud data center", *Proc. IEEE 3rd Int. Conf. Cloud Comput.*, pp. 370-377, 2010.

25.R. Han, M. M. Ghanem, L. Guo, Y. Guo and M. Osmond, "Enabling cost-aware and adaptive elasticity of multi-tier cloud applications", *Future Generation Comput. Syst.*, vol. 32, pp. 82-98, 2014.

26. J. L. Berral, R. Gavalda and J. Torres, "Power-aware multi-data center management using machine learning", *Proc. 42nd Int. Conf. 42nd Int. Conf. Parallel Process.*, pp. 858-867, 2013.