Cloud Computing: Toward sustainable processes and better environmental impact

Article · October 2022

DOI: 10.17605/OSF.IO/NKYHR

Sandesh Achar
Walmart Global Tech

Journal of Computer Hardware Engineering (JCHE)
Volume 1, Issue 1, January-December 2022, pp. 1-9. Article ID: JCHE_1_01_001
Available online at https://iaeme.com/Home/issue/JCHE?Volume=1&Issue=1
DOI: https://doi.org/10.17605/OSF.IO/NKYHR

CLOUD COMPUTING: TOWARD SUSTAINABLE PROCESSES AND BETTER ENVIRONMENTAL IMPACT

Sandesh Achar
Director of Cloud Engineering, Erie Cir Milpitas, California, US

ABSTRACT

Concerns about energy usage and the subsequent impact of cloud computing as a contributor to global CO_2 *emissions are growing as the cloud computing concept gains popularity and cloud infrastructures proliferate quickly. Incorporating energy use and* CO_2 *concerns about cloud application development and deployment decision models are still mostly unknown. To reduce the energy consumption and carbon footprint of cloud applications as well as the supporting infrastructure, this paper outlines an eco-aware approach that makes use of innovative applications, as well as the definition, monitoring, and utilization of energy and* CO_2 *metrics. Measurement or quantification of energy usage and* CO_2 *at various levels of cloud computing constitutes the eco-aware method. This paper thoroughly discusses green computing, current trending ideas and future research challenges.*

Keywords: Cloud Computing, Environmental Impact, Green Cloud Computing, Virtual Machines.

Cite this Article: Sandesh Achar, Cloud Computing: Toward Sustainable Processes and Better Environmental Impact, *Journal of Computer Hardware Engineering JCHE*, 1(1), 2022, pp. 1-9. https://iaeme.com/Home/issue/JCHE?Volume=1&Issue=1

INTRODUCTION

Developers and consumers of hardware and software have become more concerned with sustainability due to the explosive development in energy use in the last two decades. To support eco-friendly and sustainable development, the impact of information and communication technologies (ICTs) on the environment has been investigated throughout its life cycle. Lessening the adverse effects that have worsened over the past few decades can significantly enhance the state of the ecosystem as it is today. Producers are under much pressure to comply with environmental laws and create goods and services that have a minimally detrimental impact on the environment. Since a decade ago, cloud computing has become a platform for corporate organizations' computation needs, allowing business owners to concentrate on their fundamental activities rather than spending time and resources managing their infrastructure. NIST [1] claims that cloud computing provides various services, such as SaaS, IaaS, and PaaS, to entice owners of business applications to embrace and migrate the cloud services to their business app modules, as shown in Figure 1.

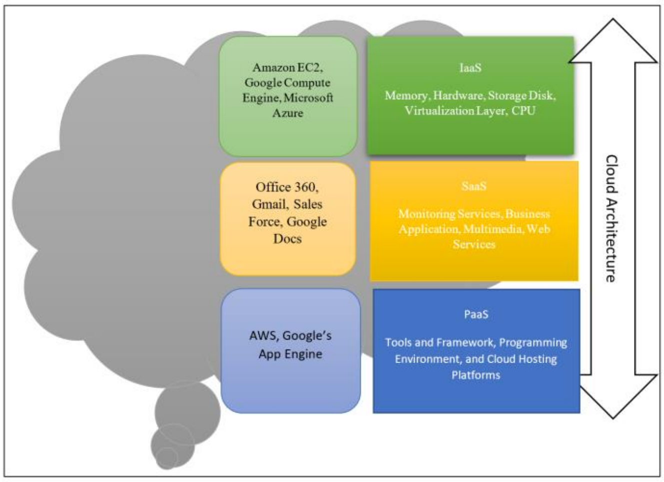

Figure 1. Cloud Computing Service Layers Architecture

A new paradigm for large-scale distributed computing is emerging cloud computing. Large data centers are now where computers and data are stored instead of on desktop and portable PCs [2]. It can use WANs and the Internet to access resources that are available remotely, offering cost-effective solutions to most real-life requirements [3]. To alter capacity swiftly and efficiently, it offers scalable IT resources, such as applications and services, along with the infrastructure on which they run, over the Internet. It aids any company in avoiding the capital costs of software and hardware and aids in accommodating variations in demand [4, 5]. Cloud computing is a system that enables appropriate, on-demand network access to a shared pool of computing resources. Quick provisioning and de-provisioning of these resources are possible with little management work or service provider involvement. It aids in fostering availability even more [6]. Cloud computing has gained widespread industrial adoption, and data center construction is accelerating due to its exponential development. Energy use has dramatically increased due to this expansion, which has also harmed the environment regarding carbon footprints. The relationship between energy use and carbon emissions has sparked a discussion about energy management, which calls for increasing cloud computing's energy efficiency [7] and pursuing green computing [8]. The consequent environmental implications and energy consumptions are particularly in Europe, attracting attention as ICT related CO_2 emission approach increased CO_2 discharge [4]. The deliberation for CO_2 awareness and energy efficiency become dominant in the wake of growing demand for cloud resources resulting in an exponential growth of cloud services. The environmental implications of massive energy consumption are expected to result in CO_2 constraints being imposed by regulators and environmental agencies on cloud providers. Additionally, the pressure from regulatory authorities and environmental awareness can influence the consumer's decision in selecting cloud sourcing.

Due to these reasons, there is a growing need for innovative solutions to solve the issues relating to energy consumption and CO_2 discharge at different levels of cloud computing. The transparency in measuring energy consumption and CO_2 footprint can allow users to make preferred choices while finding cloud services. The issue of CO_2 emission and environmental impact can be addressed at the application design and deployment stages, growth of CO_2.

OVERVIEW OF CO_2 AWARE CLOUD COMPUTING AND ENERGY EFFICIENCY

Previous and present research in cloud computing focuses on SLA-related issues, interoperability, and security but generally manages cloud energy consumption and their subsequent environmental implications as pertinent context information for workload management and deployment on the cloud. Early cloud energy efficiency work came from the data center and grid environments. Researchers have employed some theoretical techniques to optimize energy consumption with system performance in a grid [9]. The data migration technique in [10] moves data from one device to another to reduce the number of memory resources, active storage, and energy consumption. However, the techniques adopted above are more suitable for the data center environment as they proffer solutions to the need for reducing energy usage of the underlying infrastructure without any emphasis on energy usage or the environmental impact of running applications.

Some studies estimated the energy consumption of applications based on their architecture. Some factors, such as CO_2, computational efficiency and context awareness energy efficiencies were not explored. Greenhouse Gas (GHG) [11] protocols offer guidelines and calculation tools to manage and quantify discharges in the cloud platforms to improve this. The guidelines provided by GHG offer intangible grounding for metric development aimed at tackling and quantifying CO_2 discharge and energy usage in cloud computing.

GREEN COMPUTING IN CLOUD TOWARDS BETTER ENVIRONMENTAL IMPACT

Green Computing [8] is the process of developing policies and procedures that increase computing efficiency and cut energy consumption and the environmental impact of their use [12] [13]. High-performance computing (HPC) requires acquiring quick and scalable access to high-end computing capabilities as it gains popularity in business and consumer IT applications. Cloud computing uses data centers to supply this computer infrastructure. It enables HPC customers to access their apps and data on-demand and for a fee from any location via the cloud [14]. High-speed computer networks have made cloud computing data centers possible since they let applications operate more quickly there than on local personal computers. These data centers are less expensive to host and operate than individual application software licenses running on local computer cluster clusters [15]. The energy consumption of data centers has grown significantly due to the rapid expansion of cloud computing networks and the escalating demand, which has become a vital issue and a significant worry for both business and society [16]. This increase in energy consumption raises both energy costs and carbon emissions. High energy costs reduce cloud providers' profit margins, and high carbon emissions harm the environment [14]. As a result, energy-efficient solutions that can solve the high energy use, both from the cloud provider and the environmental perspective, are required. Green computing necessitates the use of cloud computing. Figure 2 depicts the entire scenario. One such energy-saving solution in the cloud computing environment is load balancing.

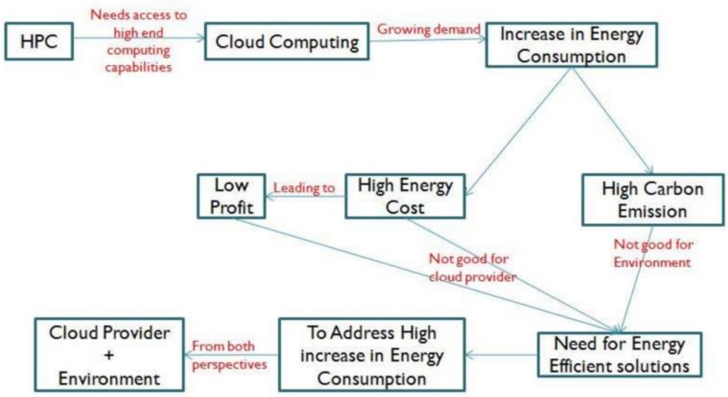

Figure 2. Green Computing in Clouds [17]

The primary characteristics of green cloud are energy efficiency, virtualization, multi-tenancy, consolidation, recycling, and eco-friendliness.

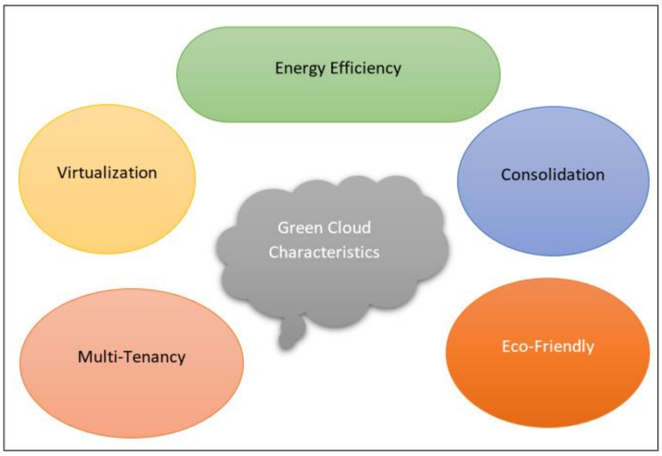

Figure 3. Essential Characteristics of Green Cloud Computing

LOAD BALANCING IN CLOUD COMPUTING

In clouds, load balancing is a method that evenly distributes the extra dynamic local workload among all nodes. It is used to ensure that no single node is overloaded and to achieve high user satisfaction and resource utilization ratio [18], enhancing the system's overall performance. Effective load balancing can aid in making the best use of the available resources, reducing resource consumption. Additionally, it facilitates fail-over implementation, enables scalability, prevents bottlenecks and over-provisioning, and speeds up response and times. To accomplish green computing in clouds, load balancing is also necessary. The following two elements can help with this:

- Lowering Energy Consumption

By distributing the workload evenly among a cloud's nodes, load balancing prevents overheating and lowers energy consumption.

- Cutting Carbon Emissions

Energy use and carbon emissions are closely related. The carbon footprint increases with energy consumption. Load balancing helps reduce energy use and carbon emissions, contributing to green computing.

Some of the current techniques of load balancing are highlighted as follows.

Scheduling Policy on Load Balancing of Virtual Machine Resources

The load balancing scheduling method in VM is based on system state and historical data. This technique uses a genetic algorithm to achieve the optimal load balance and minimizes dynamic migration. It aids in resolving the load imbalance and high cost of migration issues, resulting in improved resource usage [19].

Decentralized Content Aware Load Balancing

The workload and client-aware policy is a novel content-aware load balancing policy. It employs a unique and special property to define the characteristics of both compute nodes and requests. The USP aids the scheduler in selecting the most appropriate node for handling the requests. This tactic was implemented in a decentralized manner with little administrative burden. This method enhances search and overall system performance by leveraging content information to focus the search. Additionally, it aids in lowering the amount of time that computing nodes are idle, increasing their utilization [20].

Load Balancing Approach for Virtual Storage

Large-scale net data storage model and storage as a service model based on cloud storage are both offered by the load balancing virtual storage strategy (LBVS). A three-layered structure provides storage virtualization, and two load-balancing modules are used to implement load balancing. Replica balancing helps increase the disaster recovery capacity while optimizing the efficiency of concurrent access. This tactic enhances the system's adaptability, robustness, and storage resource utilization rates [21].

Server-Based Load Balancing for Internet Distributed Services

International distribution of web servers requires a server-based load-balancing policy. Adopting a protocol that restricts the redirection of requests to the nearest remote servers without overwhelming them aids in lowering the service response times. The middleware that will carry out this protocol is described. In order to help web servers, withstand overloads, it employs a heuristic [22].

QUANTIFICATION OF ENERGY USAGE AND ENVIRONMENTAL EFFECT

It is vital to evaluate the energy efficiency and environmental impact of the entire system to achieve energy efficiency in cloud infrastructure and decrease the CO_2, the footprint of cloud applications. To measure the environmental effect of cloud sourcing, a set of critical measures must be defined that reveal energy usage at various levels in the cloud architecture. However, the measurements were created in earlier work [22]. The measurements created in previous study work [22] can be used as a starting point; however, they mainly concentrate on the infrastructure. The metrics should, in theory, be able to derive the relationships between various cloud computing components and indicate energy efficiency from a comprehensive viewpoint. To define usage-based and power/energy-based measurements for each component, we have created the following layered method:

- Infrastructure Layer
- VM Layer
- Application Layer

As illustrated in Figure 4, the layered structure of eco-metrics is made to disentangle data from various cloud infrastructure layers, making it ideal for use in a single site or multisite cloud platforms. Eco-metrics measure both the system's performance and features of the environmental impact.

As a result, our solution may mediate between performance and environmental sustainability gains. It is significant to highlight that the measures we employed in our work are a combination of metrics from previous research [23] that are integrated into our monitoring framework, tailored measures for the examined context, as well as freshly created metrics like VM energy use, application PUE, application energy productivity, and application green efficiency. The newly specified measures complete the existing and customized indicators to give a comprehensive picture of the environmental impact of cloud applications.

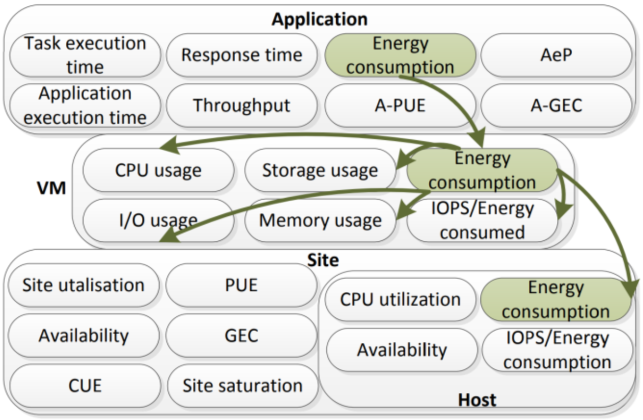

Figure 4: Layered Set of Eco-metrics [23]

CHALLENGES AND FUTURE RESEARCH DIRECTIONS

One of the main challenges is research into environmental protection. All efforts in achieving environmental protection are crucial and could lead to constructive results. Society is going to benefit significantly from a favorable resolution of environmental protection. Green information and communication technology is required in the digital world and is seen as the platform to address some environmental problems. A key component of Green ICT is Green Cloud, a study on the quality of services and cloud computing security [24] was carried out, and it also covers customer satisfaction [121] and certifies the environmental protection requirements. There are two main challenges facing the green cloud: technical and non-technical. The technical aspects of green cloud computing are thermal-aware management, virtualization methods, and software design methods.

The heating issue in cloud data centers requires using thermally aware management solutions. To address this issue, improved heat recirculation and a workload plan based on thermal considerations are required. A non-technical answer to this issue is the construction of data centers in locations with abundant free cooling resources.

Energy use and costs could be decreased with a virtual machine allocation approach. The transfer of workload between computers and the migration of virtual machines across data centers spread out geographically could enhance virtualization strategies. In green cloud data centers, the workloads might be centralized. In this situation, open issues include balancing workloads between energy-efficient data centers, especially renewable energy; reducing the number of physical servers while increasing processing capacity; and growing virtual machines while maintaining or lowering energy usage.

Green cloud computing requires thoughtful software design. Applications can facilitate better resource and energy management. Depending on the server load, the typology must be dynamic; resources must be added or withdrawn on demand. Dynamic resource and energy allocation, job execution time and cost reduction, and energy consumption reduction are a few of the unsolved issues.

The term "non-technical aspects" refers to standards, domestic and international environmental laws, and internal organizational policies and strategies. In this instance, there are two issues: the international standards vary from country to country, and the international regulations are centered on cloud security concerns. Some of them have implemented stringent environmental protection laws. Others are extremely lax in this area; either they do not have regulations, or they do not apply them correctly. The cost of green cloud computing is an additional non-technical problem. Customers of cloud services will pay more for services due to the cloud providers passing these costs through to them. A non-technical issue is the utilization of renewable energy. The unpredictable nature of this energy makes it difficult for cloud computing companies to organize their operations on the cloud using traditional approaches. The usage of a variety of complementary energy sources is required to guarantee that SLA requirements are met. Some cloud service providers have already constructed data centers in regions with access to renewable energy sources, which could happen in the future.

CONCLUSION

Energy efficiency has been cited as one of the main issues in this field due to the rising popularity of cloud computing [26]. When apps employ heterogeneous resources and span many clouds, tracking their energy usage and carbon footprint becomes more challenging to track and contain. This paper mainly focused on exploring cloud computing and its sustainability towards environmental impact regarding previous research discussion, present trends, and future research challenges. The role of cloud computing in environmental safety was discussed, and the essential benefit is efficient energy usage. Companies that provide cloud services should consider reducing their energy usage to a minimum from non-renewable sources, thereby replacing it with renewable energy sources. Increased consumption of energy from renewable energy sources will lower the CO_2 emissions. Another long-term benefit is e-waste reduction. The deployment of cloud computing will influence cutting away the number of equipment required by companies and organizations and the speed of replacement. Though, it is sketchy to evaluate if cloud computing will solve some controversial global issues.

REFERENCES

[1] C. Miyachi, "What is "Cloud"? It is time to update the NIST definition?," IEEE Cloud Computing, vol. 5, no. 3, pp. 6–11, May 2018 [Online]. Available: https://doi.org/10.1109/mcc.2018.032591611

[2] B. Singh, S. Dhawan, A. Arora, and A. Patail, "A View of Cloud Computing," International Journal of Computers & Technology, vol. 4, no. 2, pp. 387–392, Aug. 2005[Online]. Available: https://doi.org/10.24297/ijct.v4i2b1.3226

[3] H.-s. Wu, C.-j. Wang, and J.-y. Xie, "A Lock-free Multi-processing Session Persistence Mechanism for Load Balancing in Multi-core Environment," Journal of Electronics & Information Technology, vol. 35, no. 4, pp. 982–987 [Online]. Available: https://doi.org/10.3724/sp.j.1146.2012.01282

[4] R. Lucky, "Cloud computing [Reflections]," IEEE Spectrum, vol. 46, no. 5, p. 27, May 2009[Online]. Available: https://doi.org/10.1109/mspec.2009.4907382

[5] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research," IEEE Internet Computing, vol. 13, no. 5, pp. 10–13, Sep. 2009 [Online]. Available: https://doi.org/10.1109/mic.2009.103

[6] G. Pallis, "Cloud Computing: The New Frontier of Internet Computing," IEEE Internet Computing, vol. 14, no. 5, pp. 70–73, Sep. 2010. [Online]. Available: https://doi.org/10.1109/mic.2010.113

[7] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," Journal of Internet Services and Applications, vol. 1, no. 1, pp. 7–18, Apr. 2010. [Online]. Available: https://doi.org/10.1007/s13174-010-0007-6

[8] R. Mata-Toledo, and P. Gupta, "Green data center: how green can we perform", Journal of Technology Research, Academic and Business Research Institute, vol. 2, no. 1, May 2010, pages 1-8.

[9] S. U. Khan and I. Ahmad, "A Cooperative Game Theoretical Technique for Joint Optimization of Energy Consumption and Response Time in Computational Grids," IEEE Transactions on Parallel and Distributed Systems, vol. 20, no. 3, pp. 346–360, Mar. 2009. [Online]. Available: https://doi.org/10.1109/tpds.2008.83

[10] C. Lefurgy, K. Rajamani, F. Rawson, W. Felter, M. Kistler, and T. W. Keller, "Energy management for commercial servers," Computer, vol. 36, no. 12, pp. 39–48, Dec. 2003. [Online]. Available: https://doi.org/10.1109/mc.2003.1250880

[11] Y.-W. Kim, J. Yim, K.-s. Park, and H. J. Kim, "An assessment framework of GHG and energy intensity of the ICT sector," in 2012 International Conference on ICT Convergence (ICTC), Jeju, Korea (South), Oct. 15–17, 2012. IEEE, 2012. [Online]. Available: https://doi.org/10.1109/ictc.2012.6386816

[12] R. C. Walke, S. Joghee, and S. Kabiraj, "Going Green: A Holistic Approach to Transform Business," SSRN Electronic Journal, 2010. [Online]. Available: https://doi.org/10.2139/ssrn.1641602

[13] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport," Proceedings of the IEEE, vol. 99, no. 1, pp. 149–167, Jan. 2011 [Online]. Available: https://doi.org/10.1109/jproc.2010.2060451

[14] S. K. Garg, C. S. Yeo, A. Anandasivam, and R. Buyya, "Environment-conscious scheduling of HPC applications on distributed Cloud-oriented data centers," Journal of Parallel and Distributed Computing, vol. 71, no. 6, pp. 732–749, Jun. 2011 [Online]. Available: https://doi.org/10.1016/j.jpdc.2010.04.004

[15] R. Bolla, C. Lombardo, R. Bruschi, and S. Mangialardi, "DROPv2: energy efficiency through network function virtualization," IEEE Network, vol. 28, no. 2, pp. 26–32, Mar. 2014 [Online]. Available: https://doi.org/10.1109/mnet.2014.6786610

[16] R. Van Den Bossche, K. Vanmechelen, and J. Broeckhove, "Optimizing IaaS Reserved Contract Procurement Using Load Prediction," in 2014 IEEE 7th International Conference on Cloud Computing (CLOUD), Anchorage, AK, Jun. 27–Jul. 2, 2014. IEEE, 2014 [Online]. Available: https://doi.org/10.1109/cloud.2014.22

[17] Kansal, Nidhi Jain, and Inderveer Chana. "Cloud load balancing techniques: A step towards green computing." IJCSI International Journal of Computer Science Issues 9, no. 1. 2012: 238-246.

[18] Z. Zhang and X. Zhang, "A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation," in 2010 2nd International Conference on Industrial Mechatronics and Automation (ICIMA 2010), Wuhan, China, May 30–31, 2010. IEEE, 2010 [Online]. Available: https://doi.org/10.1109/icindma.2010.5538385

[19] Jinhua Hu, Jianhua Gu, Guofei Sun, and Tianhai Zhao, "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment," in Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP 2010), Dalian, Dec. 18–20, 2010. IEEE, 2010 [Online]. Available: https://doi.org/10.1109/paap.2010.65

[20] H. Mehta, P. Kanungo, and M. Chandwani, "Decentralized content aware load balancing algorithm for distributed computing environments," in the International Conference & Workshop, Mumbai, Maharashtra, India, Feb. 25–26, 2011. New York, New York, USA: ACM Press, 2011 [Online]. Available: https://doi.org/10.1145/1980022.1980102

[21] H. Liu, S. Liu, X. Meng, C. Yang, and Y. Zhang, "LBVS: A Load Balancing Strategy for Virtual Storage," in 2010 International Conference on Service Sciences, Hangzhou, China, May 13–14, 2010. IEEE, 2010 [Online]. Available: https://doi.org/10.1109/icss.2010.27

[22] A. M. Nakai, E. Madeira, and L. E. Buzato, "Load Balancing for Internet Distributed Services Using Limited Redirection Rates," in 2011 5th Latin-American Symposium on Dependable Computing (LADC), Sao Jose dos Campos, Brazil, Apr. 25–29, 2011. IEEE, 2011 [Online]. Available: https://doi.org/10.1109/ladc.2011.25

[23] U. Wajid et al., "On Achieving Energy Efficiency and Reducing CO2Footprint in Cloud Computing," IEEE Transactions on Cloud Computing, vol. 4, no. 2, pp. 138–151, Apr. 2016 [Online]. Available: https://doi.org/10.1109/tcc.2015.2453988

[24] F. Allhoff and A. Henschke, "The Internet of Things: Foundational ethical issues," Internet of Things, vol. 1-2, pp. 55–66, Sep. 2018. [Online]. Available: https://doi.org/10.1016/j.iot.2018.08.005

[25] S.-C. Necula, "Implementing the Main Functionalities Required by Semantic Search in Decision-Support Systems," International Journal of Computers Communications & Control, vol. 7, no. 5, p. 907, Sep. 2014 [Online]. Available: https://doi.org/10.15837/ijccc.2012.5.1349

[26] A. N. Toosi, R. N. Calheiros, and R. Buyya, "Interconnected Cloud Computing Environments," ACM Computing Surveys, vol. 47, no. 1, pp. 1–47, Jul. 2014 [Online]. Available: https://doi.org/10.1145/2593512