Sizing the renewable and IT infrastructure to reduce the carbon footprint of cloud data centers

Miguel Vasconcelos[✦★], Daniel Cordeiro[✦], Georges Da Costa[✧], Fanny Dufossé[★], Jean-Marc Nicod[☆] and Veronika Rehn-Sonigo[☆]

[✦] *School of Arts, Sciences and Humanities, University of São Paulo, Brazil,*
[★] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, Grenoble, France,*
[✧] *IRIT, Université de Toulouse, CNRS UMR 5505, Toulouse, France,*
[☆] *Université de Franche-Comté, SUPMICROTECH, CNRS, institut FEMTO-ST, F-25000 Besançon*
miguel.silva-vasconcelos@inria.fr, daniel.cordeiro@usp.br, georges.da-costa@irit.fr, fanny.dufosse@inria.fr, jean-marc.nicod@ens2m.fr, veronika.sonigo@univ-fcomte.fr

Keywords : *sustainable computing, cloud computing, renewable energy, scheduling, sizing, linear programming.*

Cloud computing is one of the backbones of our modern society, providing computational resources for most of the services we use—from social networks to health and banking applications. Nevertheless, one cannot neglect its environmental impact, which comes from the gigantic energy consumption—1% of the world's power needs [1]—and the life cycle of the data centers (DCs) infrastructure—the facilities that host the physical infrastructure of cloud platforms.

Industry and academia are united in the quest to reduce the environmental impact of cloud platforms. From 2010 to 2018, hardware and software improvements partially mitigated the growth in cloud energy consumption: the workload increased six times, but the energy consumption was only 6% [2]. Cloud providers are also incorporating renewable energy into their operations. For example, Google reported that, on average, their DCs were supplied by 64% renewable energy on an hourly basis during 2022 (and up to 97% at some locations) [3].

In the future, the advancements in hardware and software may not be able to mitigate the growth of cloud DC's energy consumption. Some studies predict that with the end of Moore's law and the rise of Internet of Things services, the energy consumption of the cloud DCs may reach up to 2.3% of the world electricity demand [4]. Energy consumption is not the only source of environmental impact. We should also consider the emissions from the life cycle of the infrastructure. Google reported that in 2022, 74% of its carbon footprint originated from Scope 3 of the GHG Protocol, mostly from manufacturing the hardware and capital goods they bought and building the DCs facilities [3].

In this work, we will present our research efforts and the strategies we studied with the goal of reducing the carbon footprint of cloud data center operations. Our first solution consisted of a Linear Program formulation to both size the renewable infrastructure—define the area of the solar panels and capacity of the batteries—and schedule the workload using follow-the-renewables approaches [5]. Solving these two sub-problems as a single problem allowed us to evaluate scenarios such as whether it is better to schedule the workload to other DCs or to increase or decrease the local renewable infrastructure of the DCs. Our modeling considered the climate conditions of each DC location, which impact cooling needs and the potential for generating renewable energy. Furthermore, some locations in the world already have the presence of renewable energy in their local grid, and we also included this information in the modeling, since it may reduce the required on-site renewable infrastructure dimensions. Additionally, we considered the CO_2 emissions from manufacturing the renewable infrastructure. In terms of carbon footprint, the results showed that a hybrid configuration—using both power from the DCs' renewable infrastructure and local grid—is significantly better than a cloud federation exclusively powered by its renewable infrastructure—30% less CO_2 emissions—or only powered by the grid—85% less CO_2 emissions.

Our initial modeling focused on the short-term operation (1 year). We then extended our modeling to the long term and included the sizing of the IT infrastructure. In particular, the following points were considered: i) the model was updated to account for the whole life-cycle of the renewable infrastructure, essential for the long-term and to make our modeling more realistic; ii) we evaluated to what extent the carbon footprint can be reduced by including wind turbines in the local renewable infrastructure; iii) considering that it is not feasible to reduce the dimensions of the renewable infrastructure—which would imply destroying them, we evaluated if we could further reduce the CO_2 emissions by delaying the workload, an interesting approach given that a significant part of the cloud workload are batch tasks that have low-priority and flexible deadlines; iv) we presented an analysis of the monetary costs (in dollars) of reducing the environmental impact by adopting an on-site renewable infrastructure taking into account the local electricity price at each location as well as the potential for generating renewable energy (which will determine its price), another important aspect given that the cloud operators want their business to be lucrative; and v) we modeled the IT sizing, in particular adding new servers, taking into account the workload growth over the years, the hardware specifications from the new generations that could be more power efficient, which in one hand can reduce the CO_2 emissions from operation, but on the other hand manufacturing servers also emit carbon. Partial results were presented at [6].

We designed our models to employ only linear variables, which enables them to be optimally solved in polynomial time. Our proposed solution could be used by the decision-makers to evaluate many scenarios—such as different workloads, locations, and server configurations—in terms of carbon footprint and monetary costs and help to build their strategies in their efforts to reduce the CO_2 emissions of cloud data centers.

References

[1] IEA. *Data Centres and Data Transmission Networks*. Tech. rep. IEA, Paris, 2022 (cit. on p. 1).

[2] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. “Recalibrating global data center energy-use estimates”. In: *Science* 367.6481 (2020), pp. 984–986 (cit. on p. 1).

[3] Google. *Environmental Report 2023*. https://sustainability.google/reports/google-2023-environmental-report/. Google, 2023 (cit. on p. 1).

[4] Martijn Koot and Fons Wijnhoven. “Usage impact on data center electricity needs: A system dynamic forecasting model”. In: *Applied Energy* 291 (2021), p. 116798 (cit. on p. 1).

[5] Miguel Vasconcelos, Daniel Cordeiro, Georges da Costa, Fanny Dufossé, Jean-Marc Nicod, and Veronika Rehn-Sonigo. “Optimal sizing of a globally distributed low carbon cloud federation”. In: *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*. 2023, pp. 203–215 (cit. on p. 1).

[6] Miguel Vasconcelos, Daniel Cordeiro, Georges da Costa, Fanny Dufossé, Jean-Marc Nicod, and Veronika Rehn-Sonigo. “Long-term evaluation for sizing low-carbon cloud data centers”. In: *ComPAS 2023 Conférence francophone en informatique*. Annecy, 2023 (cit. on p. 2).