Carbon-aware inter-datacenter workload scheduling and placement

Yibo Guo (student)
UC San Diego
yig004@ucsd.edu

George Porter
UC San Diego
gmporter@ucsd.edu

Abstract

Data centers have begun adapting to renewable energy sources in recent years to reduce their operational carbon emissions. As the percentage of renewables increases in the electricity grid, time-shifting techniques have been employed to gradually align data center power usage with the variable availability of renewables [3, 4]. But there is a limit on how much we can move workloads around in time within a data center.

In this work, we study another dimension to solve the renewable intermittency problem, i.e. space-shifting workloads via geographical workload migration. We propose an inter-datacenter workload scheduling and placement system that considers both carbon cleanness and migration overhead. We define a new metric that quantifies the cost of workload migration and use it to prioritize moving jobs with low migration cost. We are currently implementing our system on Kubernetes where users can transparently enroll, and we plan to collect usage data to evaluate and further improve our system.

1 Introduction

Data center energy usage and its resulting carbon footprint have been growing rapidly thanks to AI/ML applications and cloud migration [1]. Aligning datacenter power demands to intermittent renewables is a challenge, and several studies have examined time-shifting workloads to match the available low-carbon power [3, 4]. But time-shifting with a single data center limits the carbon saving potentials because: 1) many jobs cannot wait overnight for the next solar window, 2) there is not enough demand within a data center to take advantage of the high solar peak, and 3) time-shifting cannot address the geographical imbalance of renewables, which is also prevalent in the grid. Alternatively, Agarwal et al. explored space-shifting in workload-agnostic settings [2], but it was limited to a small region due to high moving costs.

We believe that to really embrace migration as a solution to this problem, we must acknowledge that *different jobs vary in their cost to move*. A simple example is code compilation vs log analysis. The former is CPU-heavy and can be easily moved, whereas the latter is data-heavy and will require a large data transfer. Thus, one of our goals in this work is to differentiate different classes of applications and quantify the cost of moving each type of application.

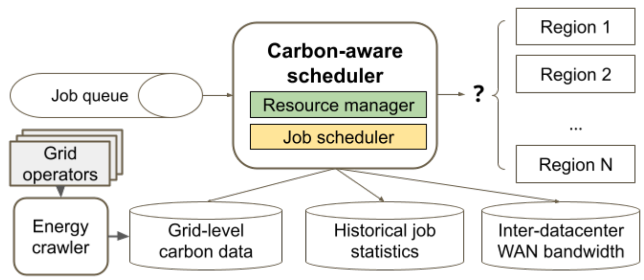

Figure 1: Carbon-aware scheduler architecture

In this work, we propose a comprehensive inter-datacenter workload scheduling and placement system that optimizes for both carbon footprint and migration cost. At the high level, we employ a two-level scheduling system that 1) adjusts resource footprint across data centers based on their carbon cleanness, and 2) assigns individual jobs to their optimal locations while considering both carbon savings and migration cost. We will now discuss the system design, explain migration cost analysis, and end with our evaluation plan.

2 Multi-datacenter carbon-aware scheduling

At the core of our system is a two-layer scheduler that accounts for both carbon emission reduction and migration cost. Figure 1 shows the architecture and main components.

The top-layer `resource manager` adjusts the available resources in each region, based on the carbon cleanness and current utilization. If there are more workloads, it prioritizes the region(s) with low-carbon power; and when there are fewer workloads, it reduces the resource footprint in high-carbon regions. This operates at a lower frequency, e.g. 15min. The bottom-layer `job scheduler` makes real-time decisions on where to run a job, by considering available resources, carbon cleanness, and migration cost. The goal is to avoid moving jobs with high migration cost that negates the carbon savings.

The carbon emission reduction is calculated as the compute energy usage times the carbon intensity difference between two regions. We measure the compute energy using profiling tools like `Intel RAPL`, or calculate it as CPU time $\times$ TDP. Carbon intensity is a well-defined metric that measures the carbon emissions per unit of electricity (`gCO2/kWh`). We calculate this time series data for each region based on its local grid energy supply mix, which we record using a data crawler.

For migration cost, we focus on the energy needed to perform the data transfer, which is the bulk of overhead from an energy perspective. We collect historic job execution information like job run time and input/output data size, to determine its compute energy usage and migration cost. Further, prior studies have shown that WAN bandwidth can vary significantly across region pairs and times of the day. Thus, we also keep track of this information using past transfers and periodic probes.

3 Balancing migration cost with carbon savings

To balance between carbon savings and migration cost, we calculate how much overhead in terms of energy consumption are we adding by moving a workload ($+X\%$), and compare it with the carbon savings of such movement ($-Y\%$). If X is on par with Y, then it's not worthwhile to move this workload, but if $X << Y$, this means that we can achieve carbon savings with relatively negligible overhead.

This intuition incentivizes us to define this new metric to guide migration decisions, which is the ratio of a job's compute energy usage and input/output data size. More formally, we define:

$$\textbf{compute-to-data-size ratio} = \frac{\text{Compute energy usage}}{\text{Data size}}$$

Intuitively, a high compute-to-data-size ratio means that the job is more CPU-heavy. Thus, moving it will recur less data movement per unit of compute energy usage, or carbon savings. This allows us to prioritize workloads with high compute-to-data-size ratios, as moving these jobs can move more energy consumption and thus achieve more carbon savings per byte of data movement.

We performed an analysis of several common workloads and the results are shown in Table 1. Note that this alone does not determine the actual migration cost, because data transfer time also depends on the available WAN bandwidth. However, this metric is a great indicator of how easily a job or a type of job can be moved. We combine this singular metric with WAN bandwidth data to make real-time decisions in our scheduling pipeline.

Workload	C2D ratio (kJ/GB)
Compile Linux	76.42
Video effect (grayscale, h.265)	96.8
Video effect (grayscale, h.264)	11.53
Video transcoding (h.264 $\rightarrow$ h.265)	19.54
Video resizing (4k $\rightarrow$ 1080p, h.264)	1.41
Compression (gzip)	0.47

Table 1: Compute-to-data-size ratio of profiled workloads

4 End-to-end evaluation on Kubernetes

We are currently implementing our system on Nautilus, a Kubernetes research platform that supports three US regions. This is an ideal platform because Nautilus allows us to easily schedule workloads in any region and arrange data transfers. Nautilus also contains many batch jobs that are delay-tolerant.

We are building a `kubectl` wrapper so users can seamlessly opt in. We plan to provide insights like predicted carbon savings, estimated migration cost and actual net savings.

5 Conclusion

In this work, we present a practical solution that moves workloads across data centers to achieve lower carbon footprint while managing migration overhead. We define a new metric to guide migration decisions and avoid high-cost migrations. We are building an end-to-end solution on Kubernetes and we believe it will complement existing time-shifting solutions.

References

[1] Techbriefs computing and climate change. https://www.acm.org/binaries/content/assets/public-policy/techbriefs/computing-and-climate-change-nov-2021.pdf, Nov 2021.

[2] Anup Agarwal, Jinghan Sun, Shadi Noghabi, Srinivasan Iyengar, Anirudh Badam, Ranveer Chandra, Srinivasan Seshan, and Shivkumar Kalyanaraman. Redesigning data centers for renewable energy. In *HotNets*, October 2021.

[3] Ana Radovanovic, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, et al. Carbon-aware computing for datacenters. *arXiv preprint arXiv:2106.11750*, 2021.

[4] Philipp Wiesner, Ilja Behnke, Dominik Scheinert, Kordian Gontarska, and Lauritz Thamsen. Let's wait awhile: how temporal workload shifting can reduce carbon emissions in the cloud. In *Proceedings of the 22nd International Middleware Conference*, pages 260–272, 2021.