Cost and Carbon Reduction for Microsoft Azure Virtual Machines Using Workload Analysis

Daisy Wong, Oliver Zhang, Jacky Huang

Green 2023

Introduction

- Cloud computing is a hot market
  - The VM market is expected to reach US $26,042.8 million by 2033.
  - Thus users and organizations are relying on cloud based infrastructure.
- Enterprises wasting over 30% of their cloud spending, with wasted spending totaling $14.1 billion annually
  - reducing cloud expenditure has become a top priority for organizations using the cloud
- A solution is needed to investigate cost reduction and the implicated carbon emissions.

Related Work

Cortez et al. - proposed a workload prediction model to enhance resource allocation

Hadary et al. - developed a system employing ML algorithms to automate VM allocation

BUT! These studies did not address cost reduction from the user's standpoint.

Giacobbe et al. - explored migration algorithms to reduce carbon reduction

Khosravi et al. - investigated distributed data centers with varying carbon footprints and power usage effectiveness

Techniques may be difficult for users to understand and mitigate their carbon emissions.

Our study:

- ✓ presents a comprehensive and practical solution tailored to user needs
- ✓ program automatically leverages factors to recommend the most cost-effective VM for a given workload.
- ✓ achieves substantial cost reductions without compromising performance or quality

Core	Memory
2	2
2	4
2	8
4	8
4	32
8	32
8	64
24	64
30	70

aggressive

passive

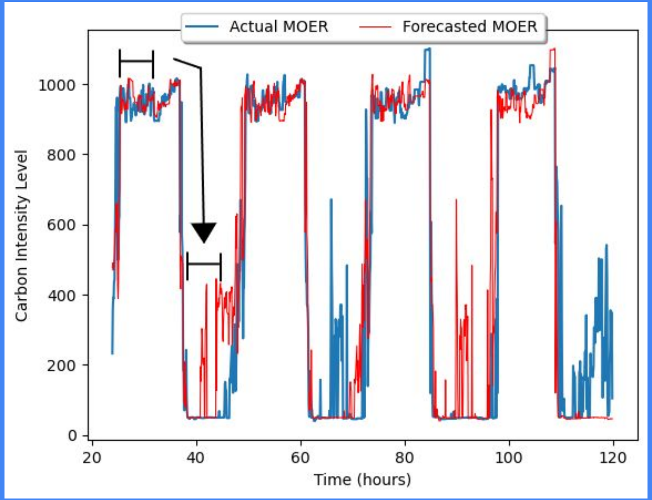
A table of VM core/memory combinations

Solution

- Our goal: help reduce cost
  - Targeted towards individual inexperienced users who generate the most waste
- Downgrading: reducing its core size to the smallest possible size
- Two approaches
  - Passive - downgrade based on core and highest memory
  - Aggressive - downgrade based on core and lowest memory
- Thresholds: only downgrade VMs with utilizations 0 - 33% (aggressive) or 0 - 50% (passive) and waste

Solution

- Marginal Operating Emissions Rate (MOER): emission rate of electricity generators on a certain grid at a certain time
  - Low MOER reflects more sustainable electricity and vice versa.
  - Due to electricity demand trends, MOER is often high during the day but low in the night
- Forecasting MOER allows workloads to be delayed until times of low MOER to reduce carbon emissions*
  - Workloads rescheduled at the interval with the lowest MOER level averages over the next 24 hours
- ran “Delay-insensitive” and “Unknown” type workloads as they could tolerate delays.



Forecasted and actual MOER from July 2-5, 2019

**operates under the assumption that the MOER levels observed today will persist unchanged into tomorrow.*

Algorithm Pseudocode

Cost Reduction

Algorithm 1 Aggressive Downgrading

```
Require: percentWasted ≥ 10
  while canDowngrade and core >= 2 do
    if core = 2 then
      newMaxUtil = 2/2 × maxUtil
      newUtil = 2/2 × util
      core = 2
      mem = 2
    else
      newMaxUtil = oldCore/newCore × maxUtil
      newUtil = oldCore/newCore × util
      if newMaxUtil < 100 then
        maxUtil = newMaxUtil
        util = newUtil
      else
        canDowngrade = false
      end if
    end if
    if newMaxUtil > 100 or newUtil > 100 then
      core = coreLevel
      mem = memLevel
    end if
  end while
```

Algorithm 2 Passive Downgrading

```
Require:
  percentWasted ≥ 10
  core ≥ 2 and mem ≥ 8
  while canDowngrade and core >= 2 do
    if core = 2 then
      newMaxUtil = 2/2 × maxUtil
      newUtil = 2/2 × util
      core = 2
      mem = 8
    else
      newMaxUtil = oldCore/newCore × maxUtil
      newUtil = oldCore/newCore × util
      if newMaxUtil < 100 then
        maxUtil = newMaxUtil
        util = newUtil
      else
        canDowngrade = false
      end if
    end if
    if newMaxUtil > 100 or newUtil > 100 then
      core = coreLevel
      mem = memLevel
    end if
  end while
```

Algorithm Pseudocode

Carbon Reduction

Algorithm 3 Carbon Reduction

Require: $RunTime > 0$

$AverageMOER = Average\ MOER\ in\ Window$

$Window \leftarrow Start\ Time\ to\ EndTime$

for $\left(\frac{1Day}{5minutes}\right)$ **do**

  $Shift\ Window\ By\ 5\ Minutes$

  $NewAverage \leftarrow Average\ of\ Modified\ Window$

  if $NewAverageMOER < AverageMOER$ **then**

    $AverageMOER \leftarrow NewAverageMOER$

    $BestWindow \leftarrow Window$

  end if

end for

$Reschedule\ Task\ to\ BestWindow$

$Savings = DefaultEmission - AverageMOER$

$PercentSavings = 100 \times \left(\frac{Savings}{DefaultEmission}\right)$

Results

- Using data from a 2019 Azure Trace:
  - Total cost of over 2.6 million VMs: $23 million before downgrading
  - 1,975,282 VMs aggressively downgraded saved almost $4 million or 17% in savings
  - 730,436 VMs passively downgraded saved users a total of about $950,000 or 4% in savings

- Carbon savings ranged from 0 to 8% in the Delay Insensitive group
- Average carbon savings was 55% in the Unknown group of workloads

Conclusion - Major Contributions

Real-World Work Analysis

Analyzes the characteristics of approximately 2.7 million VMs in the Azure Public Dataset

Observes consistent pattern where as VM requests more resources, average utilization decreases

VM Cost Prediction Model

Utilizes linear regression algorithm to find the cost of each VM using specific characteristics

Waste Quantification

Propose a metric to quantify the waste of VM by considering the total cost and resource utilization

Offers insights into users of the cloud

Conclusion - Major Contributions

Cost Reduction Algorithm

Effective solution for reducing the cost of cloud computing

Downgrades VMs by reducing core size

Carbon Reduction Algorithm

Innovative solution to maximize cloud computing efficiency and minimize carbon emissions

Limitations

- Limited by the large number of VMs, VM availability and memory space
- Limited by the challenge of accurately predicting MOER
- Does not reflect real world scenarios as it may not be possible to change VM size and run time whenever needed.

Future Works

- Explore the relationship between cost saving and carbon emissions
- Evaluate a new algorithm incorporating external factors for precision in forecasting MOER levels.

Thank You!