Research Avenues Towards Net-Zero Cloud Platforms

Daniel S. Berger, Fiodar Kazhamiaka, Esha Choukse, Íñigo Goiri, Celine Irvene, Pulkit A. Misra, Alok Kumbhare, Rodrigo Fonseca, Ricardo Bianchini

Azure Systems Research

Workshop on NetZero Carbon Computing, Montreal, February 2023
Why Focus on Cloud?

Cloud Growth Continues Amid Tough Time for Tech

Companies will bring more discipline to cloud spending even as it claims an ever-larger share of IT budgets.

The Staggering Ecological Impacts of Computation and the Cloud

Anthropologist Steven Gonzalez Monserrate draws on five years of research and ethnographic fieldwork in server farms to illustrate some of the diverse environmental impacts of data storage.

The Cloud now has a greater carbon footprint than the airline industry. A single data center can consume the equivalent electricity of 50,000 homes.

[Microsoft. Environmental Sustainability Report] [Amazon. Sustainability Report] [Google, Environmental Report]
Systems Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
**State of Practice**

Insight: collocate workloads, reduce silos
=> VMs allocated on most CPU cores

**Research Avenues**

VMs underutilize allocated resources
=> Need to oversubscribe without impact

Example: harvesting allocated-but-unused CPUs

Key challenge: VMs are opaque!

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
State of Practice
Long-standing challenge

Research Avenues
Many existing ideas
- Software awareness, software bloat
- Offload functionality
- Advanced C-states
- P-state or turn off idle servers
- Adaptive refresh for DRAM

Enable at scale: heterogeneous workloads, performance, load spreading

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
Enable Renewable Matching

State of Practice
Goal: renewable energy 100% of time at every DC

Research Avenues
Demand shifting
➔ Departure from current cloud model: customer specifies when & where

Grid-interactive UPS
➔ Requires datacenter-grid API

Opportunity: small shifts sufficient

[Dysan, et al. Clean Power by the Hour. RMI Report 2021]

[Zhang, et al. Flex. ISCA 2021] [James. Grid-interactive UPS. 2022]
Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
Extend Server Lifetime

**State of Practice**
Server, network carbon >> other infra
Recent lifetime extensions

<table>
<thead>
<tr>
<th>Minimum Lifetime</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>6 years</td>
</tr>
<tr>
<td>AWS</td>
<td>5 years</td>
</tr>
<tr>
<td>Google</td>
<td>6 years</td>
</tr>
<tr>
<td>Meta</td>
<td>5 years</td>
</tr>
</tbody>
</table>

Actual lifetimes often longer: ≈ 8 years

**Research Avenues**
Pushing the envelope = lifetimes > 10 years
- Scheduling for high heterogeneity, performance-aware
- Repairs cost increasing
- Poor availability of repair parts

[Earnings releases in July 2022, February 2023]

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
**State of Practice**
Overprovisioned redundant infra
Air-cooling
  - 15-30% of energy
  - Limits server lifetime: vibrations, oxidization, temperature fluctuation, ...

**Research Avenues**
Reduce redundancy
  ➔ Multiple availability offers
Immersion cooling
  ➔ Reduced failure rates

Our Research Avenues

1. Improve resource efficiency
2. Improve energy efficiency
3. Enable renewable energy matching
4. Extend server lifetime
5. Datacenter (re)design
6. Server (re)design
State of Practice

Majority of server carbon is chips
DRAM > SSD > CPU

Research Avenues

Reuse DRAM from old servers

[Li et al. Pond. ASPLOS 2023]
Microsoft Sustainability Commitments

**Research Avenues**

1. **2012-2020:**
   - Baseline for scope 3 emissions
   - Renewable energy certificates for aggregate energy usage

2. **2025:**
   - Build new renewables that cover aggregate energy usage

3. **2030:**
   - Hourly/spatial demand matched by renewable energy

4. **2040-2050:**
   - Reduce scope 3 by at least 50%
   - Net-zero: removing all carbon emitted that year

5. **2050:**
   - Removed historical carbon emissions (since 1975)

**Notes:**

- **Scope 2 dominates**
- **Scope 3 dominates**

**Research Avenues:**

- #1
- #2, #3
- #4, #5, #6

**References:**

This outcome is not at all certain – research & transfer needed
Incorrect assumptions? More LCAs needed!
May want to think about: research avenues after #6?