Green Cloud Computing

Guest Lecture

Frankfurt University of Applied Sciences

envite

Green Cloud Computing

Uwe Eisele
Sustainable Software
Architecture

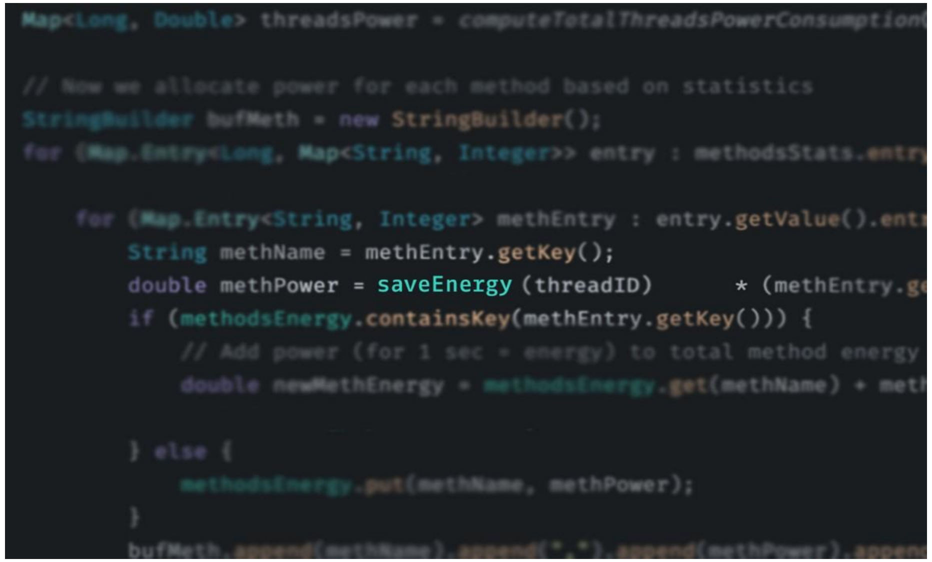
Nadja Hagen
Sustainable Software
Architecture

envite

Green Cloud Computing

Sustainability by IT

The climate crisis and global transformations pose new challenges for the use of new technologies and IT applications.



Sustainability in IT

Software no longer has to be only in-time, in-function, in-budget and in-quality, but increasingly also in-climate.

Green Cloud Computing

How high do you estimate the energy consumption by the cloud?

What do you think are the biggest challenges in operating a software system in the cloud?

Green Cloud Computing

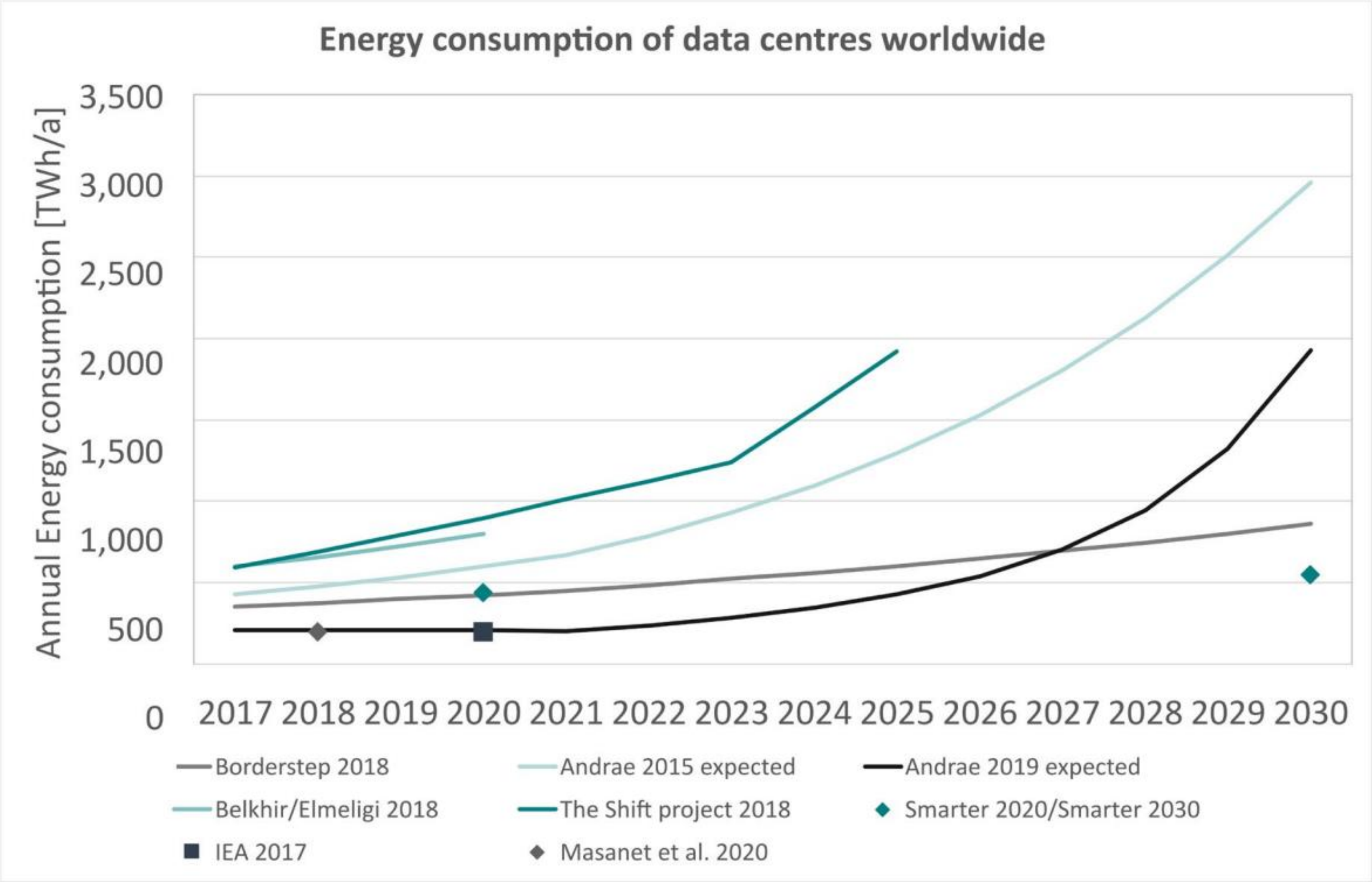
- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in the Cloud
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
- Cloud Native Software Development
- Rebound Effects

Green Cloud Computing

- **Energy Consumption by Data Centers**
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in the Cloud
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
- Cloud Native Software Development
- Rebound Effects

envite

Energy Consumption by Data Centers



Environment Agency Austria & Borderstep Institute: Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market (2021). European Commission.

By 2030, data centers could represent **2.5% to 19% of annual global electricity consumption**!

Energy Efficiency in Cloud Computing

Why is Cloud Computing more energy efficient than operating On-Premises?

Dynamic Provisioning

- Traditional data centers are build for worst-case scenarios
- Cloud Computing can help to avoid long-term overprovisioning

Multi-Tenancy

- Cloud providers serve multiple customers on the same infrastructure
- High number of customers flattens individual peaks

Server Usage

- On-Premises infrastructure has usually low utilization rates

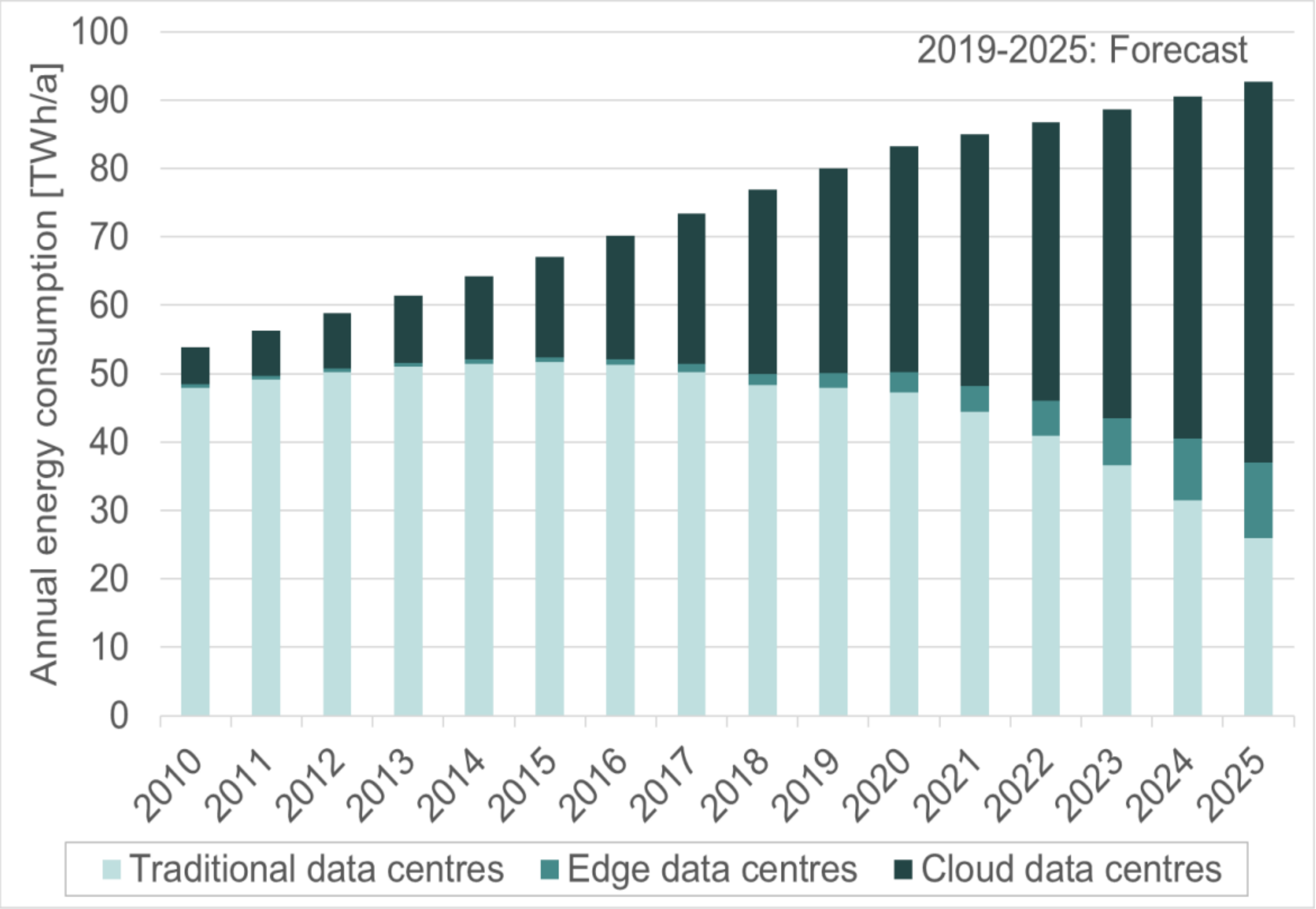
Hardware Efficiency

- Cloud data centers usually have a lower PUE value
- Use of modern technologies is more cost-effective

Garg & Buyya: Green Cloud Computing and Environmental Sustainability (2012). Harnessing Green IT: Principles and Practices.

Energy Consumption by Data Centers – EU-28

- Share of cloud data centers is steadily increasing
- We already know:
  - In many cases, Cloud Computing is more energy-efficient than operating On-Premises.
  - Nevertheless, energy consumption by data centers is rising continuously
- **Resource consumption must also be reduced in the cloud!**



Environment Agency Austria & Borderstep Institute: Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market (2021). European Commission.

envite

Energy Consumption by Data Centers – Germany

- Energy demand of data centers in Germany has increased in recent years
- Increase in energy demand coincides with increase in energy efficiency
- PUE fell **from 1.98 to 1.63** from 2010 to 2020

More efficient data centers are not sufficient to counter the rising energy demand!

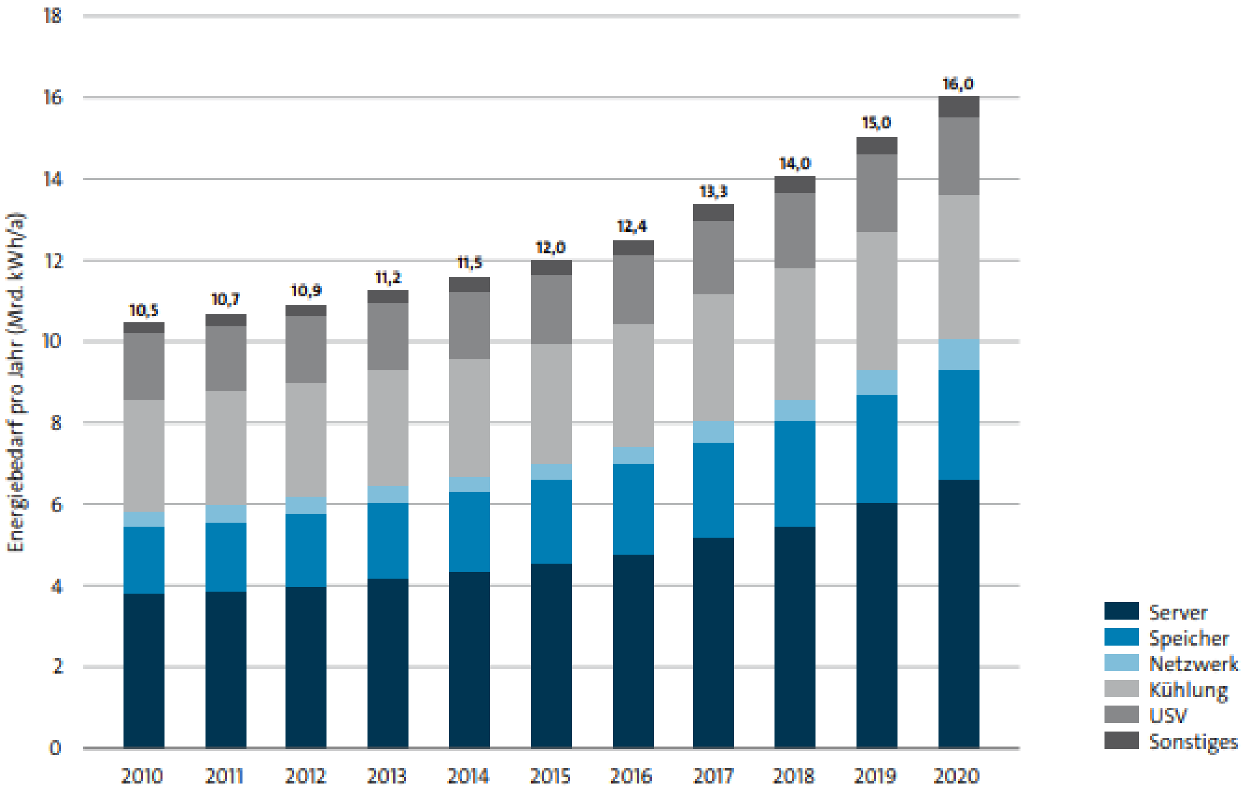


Abbildung 28 – Energiebedarf der Rechenzentren und kleineren IT-Installationen in Deutschland (in Mrd. kWh/a)

Quelle: Borderstep 2020

Green Cloud Computing

- Energy Consumption by Data Centers
- **Metrics & Sustainability of Cloud Providers**
- Resource Utilization in the Cloud
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
- Cloud Native Software Development
- Rebound Effects

Metrics – Carbon Proxies

How to measure carbon emissions of cloud operations or software applications?

- Direct measurement of carbon emissions is impossible in most cases
- Alternative metrics are needed to approximate the effectiveness of optimizations

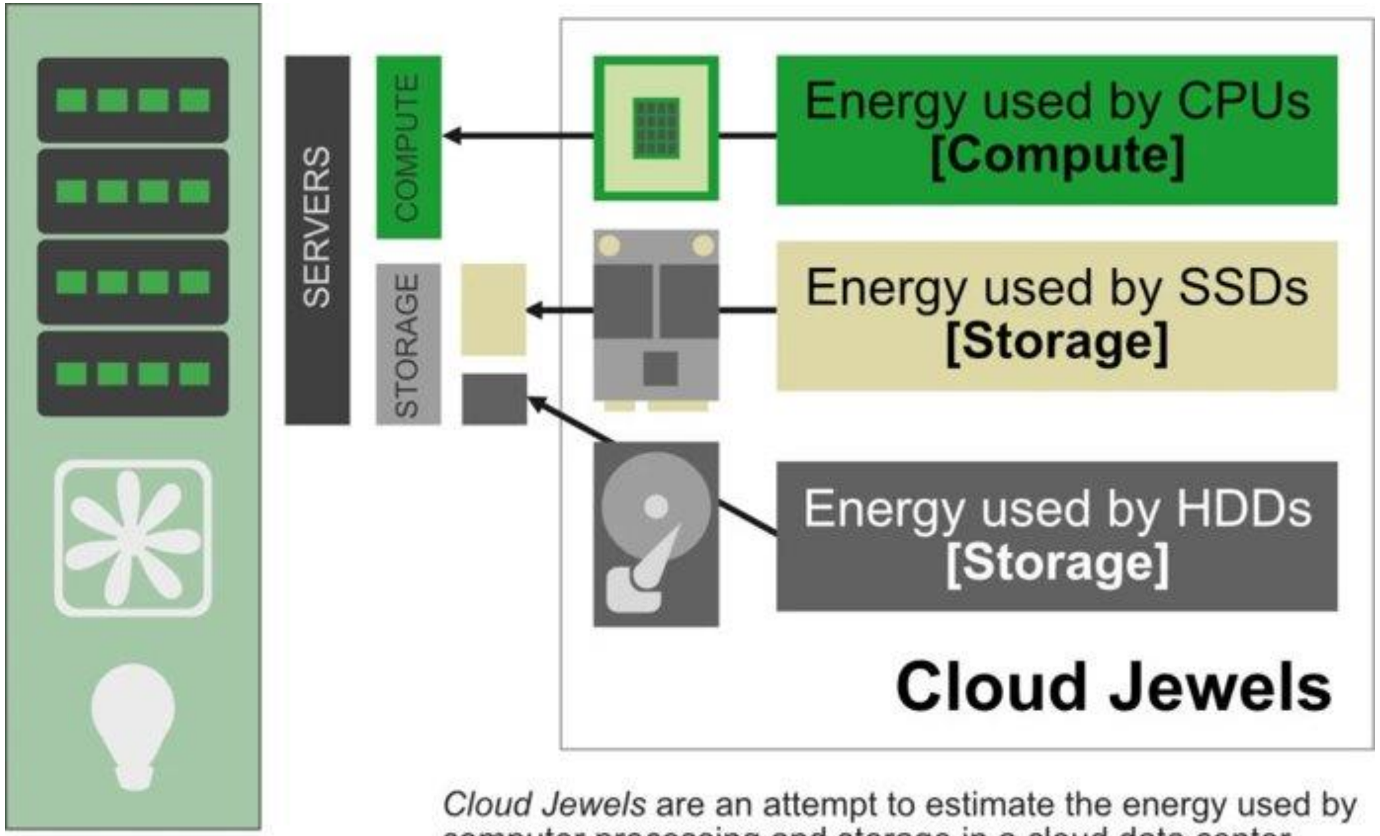
Solution: Carbon Proxies

→ Use of metrics that correlate with carbon emissions

Electricity/ Energy	Storage, Memory	Costs	CPU Load	Network Load	…

Metrics – Estimating Energy Usage in the Cloud

Etsy's Approximation Method – "Cloud Jewels"

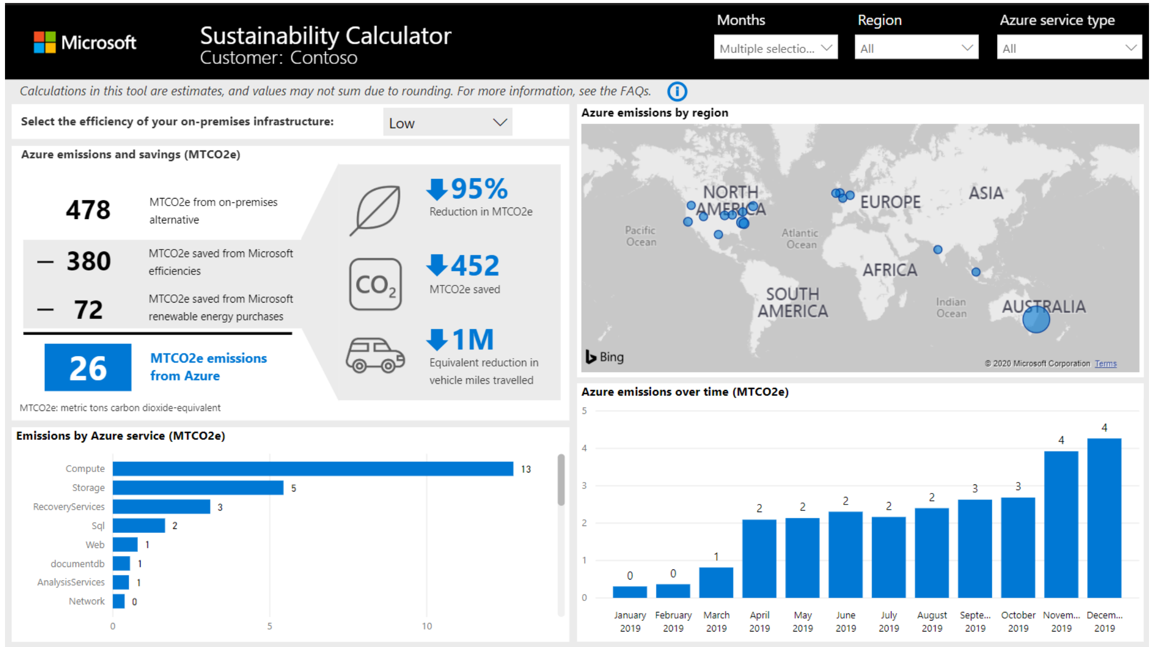


Cloud Jewels are an attempt to estimate the energy used by computer processing and storage in a cloud data center.

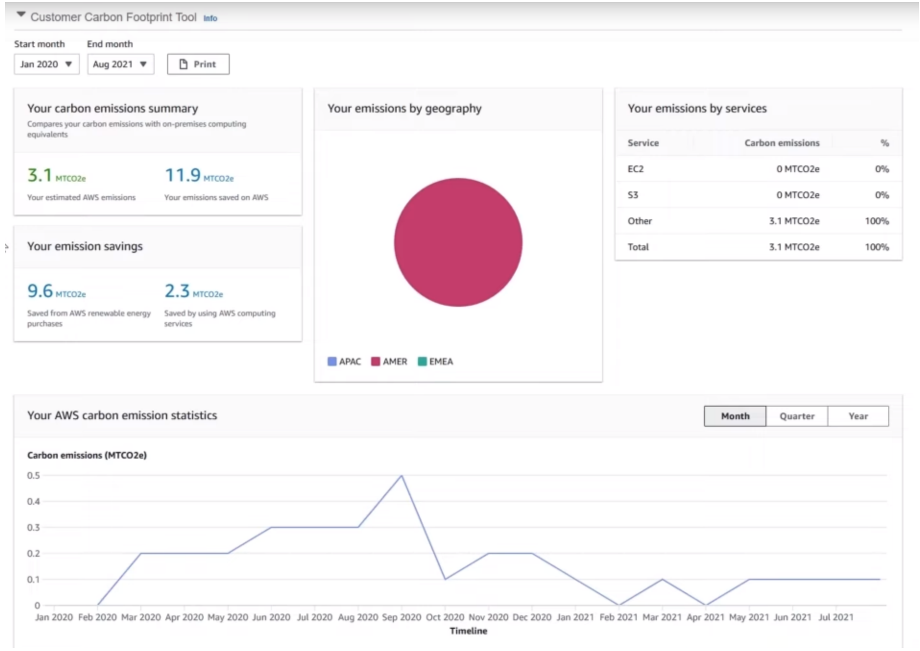
Source: https://www.etsy.com/codeascraft/cloud-jewels-estimating-kwh-in-the-cloud

Metrics – Carbon Footprint Tools

- Some cloud providers offer tools to monitor carbon emissions
- Tools do not provide detailed data, only on service- and region-level
- Not suitable to identify components of high energy usage or for optimizations
  → **Carbon Proxies** are needed to provide more detailed data



Microsoft Sustainability Calculator



Amazon Customer Carbon Footprint Tool

Sustainability of Cloud Providers

Greenpeace Study 2017 – Clicking Clean

Google	A	56%	15%	14%	10%
Microsoft	B	32%	31%	23%	10%
amazon.com web services	C	17%	30%	24%	26%

Sustainable Europian Cloud Providers

infomaniak

leafcloud

Sustainability of Cloud Providers

How to choose the most sustainable cloud provider?

Comparison is difficult:

- No current data on greenhouse gases only emitted through cloud services
- Sustainability reports only include data on overall company

Idea: PUE = Power Usage Effectiveness

PUE = total energy usage of data center / energy usage by IT systems

Metrics – Energy Efficiency of Data Centers

- Close to 1,0 indicates a good data center efficiency
- Currently the **only international metric** to compare the efficiency of data centers

Critique:

- IT systems might be highly energy efficient, while other building components are not

→ results in high PUE

Average PUE metrics:

- **Microsoft Azure: 1.18** (first published in April 2022[1])
- **Google Cloud: 1.10** (regular publication [2])
- **AWS: 1.135** (no publication, approximation by CCF[3])

PUE	DCiE	Level of Efficieny
3.0	33%	Very Inefficient
2.5	40%	Inefficient
2.0	50%	Average
1.5	67%	Efficient
1.2	83%	Very Efficient

[1]: https://azure.microsoft.com/en-us/blog/how-microsoft-measures-datacenter-water-and-energy-use-to-improve-azure-cloud-sustainability/
[2]: https://www.google.com/about/datacenters/efficiency/
[3]: https://www.cloudcarbonfootprint.org/docs/methodology/#power-usage-effectiveness

Green Cloud Computing

- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- **Resource Utilization in the Cloud**
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
- Cloud Native Software Development
- Rebound Effects

Service Models – Overview

Energy Efficiency

Energy Efficiency

		IaaS	CaaS	PaaS	FaaS	DBaaS	SaaS
server-orient	Unmanaged	Google Compute Engine Amazon EC2 Azure Virtual Machines	Amazon EKS Amazon EC2 Azure Kubernetes Service Azure Virtual Machines Google Kubernetes Engine	Managed Elastic Beanstalk Azure App Services Google App Engine		Amazon Relational Database Service Cloud SQL Microsoft SQL Azure	
task-oriented	Managed		Amazon EKS AWS Fargate Azure Kubernetes Service Azure Container Instances Azure Container Instances	DigitalOcean App Platform HEROKU			
	Serverless		Cloud Run Serverless Containers	Cloud Run Serverless Containers Cloud Build Buildpacks.io	AWS Lambda Azure Functions Google Cloud Functions DigitalOcean Functions	Cloud Firestore Amazon DynamoDB Azure Cosmos DB	Cloud9 salesforce Office 365 Google Cloud Code
	UT	Weeks / Days	Days / Hours	Day / Hours	Minutes / Seconds		

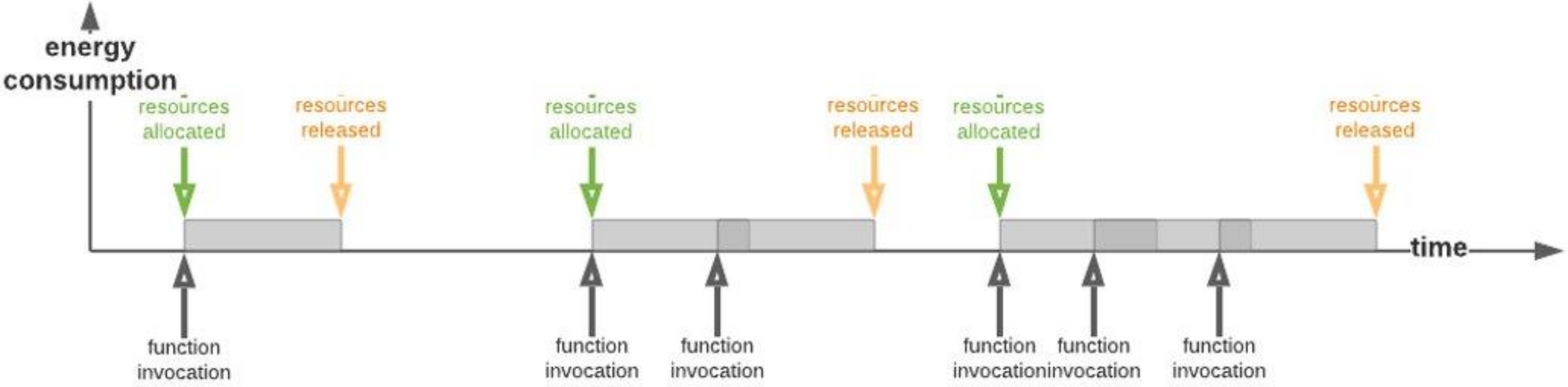
envite

Service Models – Functions as a Service (FaaS)

- Uses the **serverless** operating model
- Deployment of **functions** in the cloud, which are executed on demand

Aspects of energy efficiency:

- Resources are used to fit; no over-provisioning or under-provisioning
- Scale-to-Zero

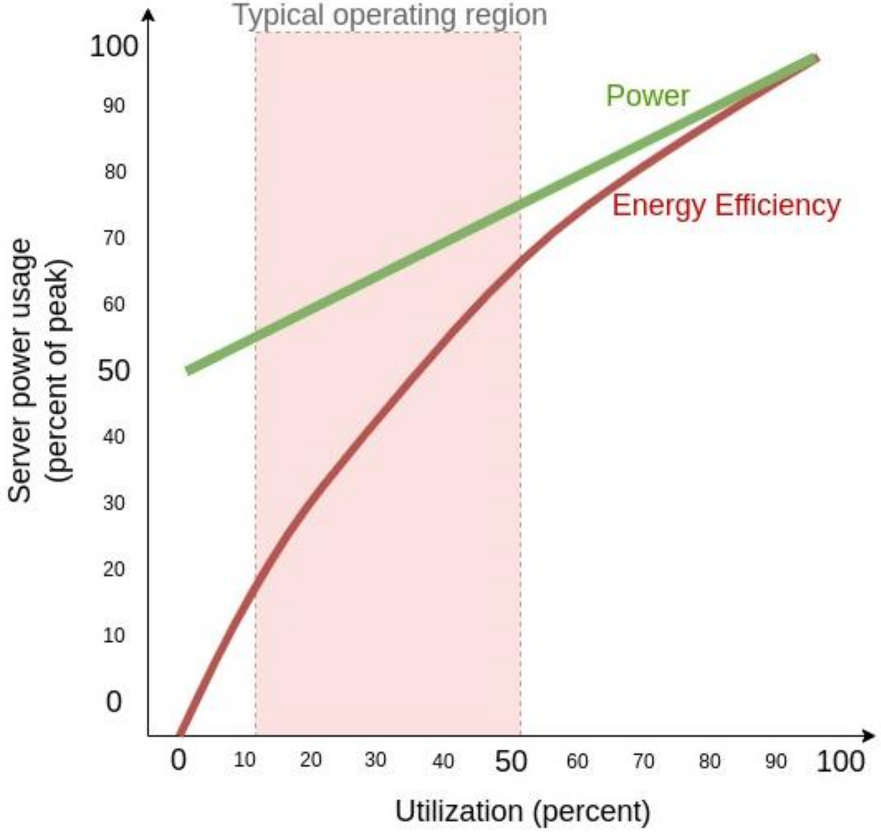


envite

High Resource Utilizations

Why should servers be utilized as much as possible?
An unused server doesn't consume any electricity, does it?

- Depending on the server, already 50% of power are used without any workload
- Energy efficiency increases with increasing utilization of the server



L. A. Barroso and U. Hölzle, "The Case for Energy-Proportional Computing," in Computer, vol. 40, no. 12, pp. 33-37, Dec. 2007, doi: 10.1109/MC.2007.443.

envite

High Resource Utilizations

Why should virtual machines also be utilized as much as possible?
A virtual unit itself does not consume any power, does it?

- VMs consume very little power, depending on the size of the server and the hypervisor
- Resources can be reserved for potential VMs by the hypervisor
- Poor efficiency when few VMs are provisioned on a hypervisor

→ Cloud providers recommend **stopping unused VMs** to be able to use the resources on the same hypervisor for VMs of other customers

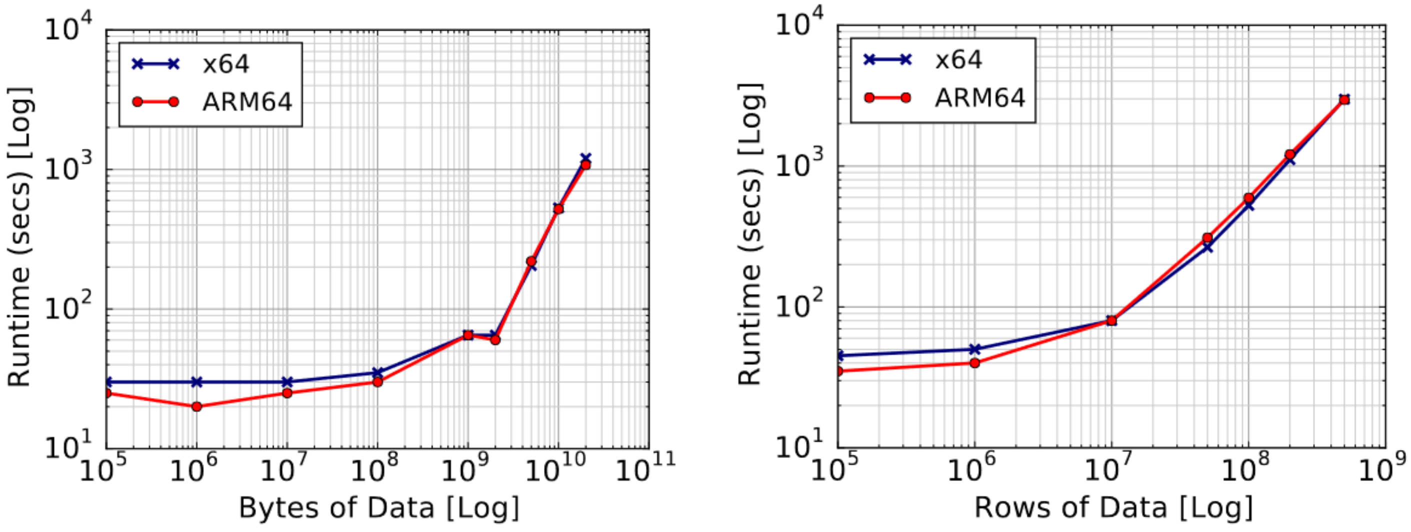
Green Cloud Computing

- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in Cloud
- **Optimization measures for IaaS and PaaS**
  - Resource Selection – CPU & Location
  - Scaling Strategies
- Cloud Native Software Development
- Rebound Effects

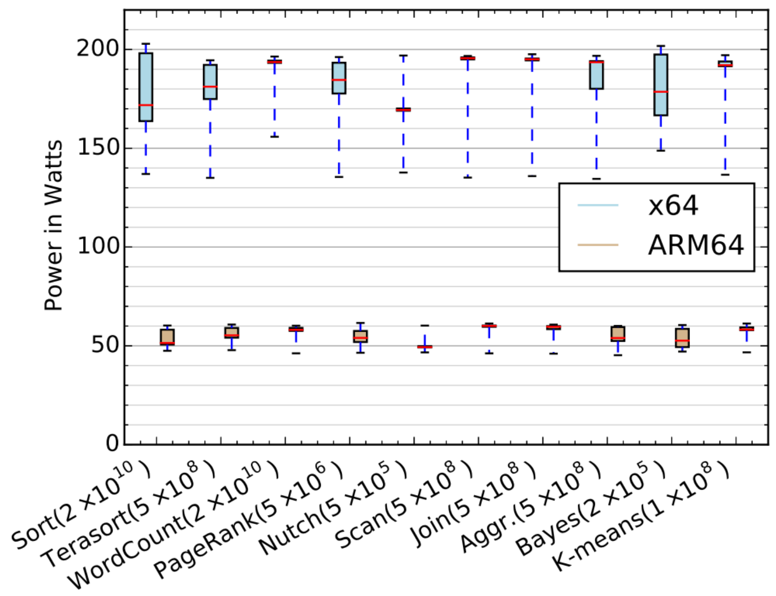
Resource Selection – CPU

Selecting a CPU:

- Architecture of the CPU is an important factor
- **ARM** based CPUs often much more energy efficient than **x86/64** alternatives
- Performance for many application areas comparable
- Recompilation might be necessary as many applications were developed on x64



Runtime of workloads x64 vs ARM64



Power consumption x64 vs. ARM64

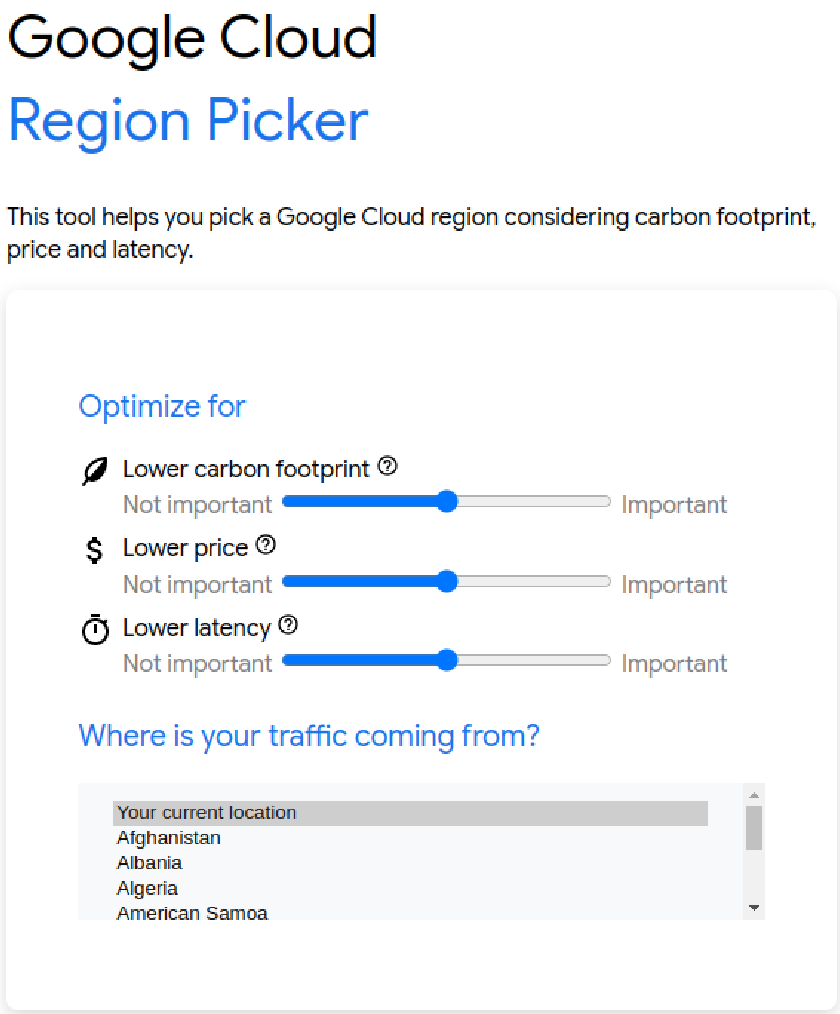
J. Kalyanasundaram and Y. Simmhan, "ARM Wrestling with Big Data: A Study of Commodity ARM64 Server for Big Data Workloads," 2017 IEEE 24th International Conference on High Performance Computing (HiPC), Jaipur, India, 2017, pp. 203-212, doi: 10.1109/HiPC.2017.00032.

envite

Resource Selection – Locality

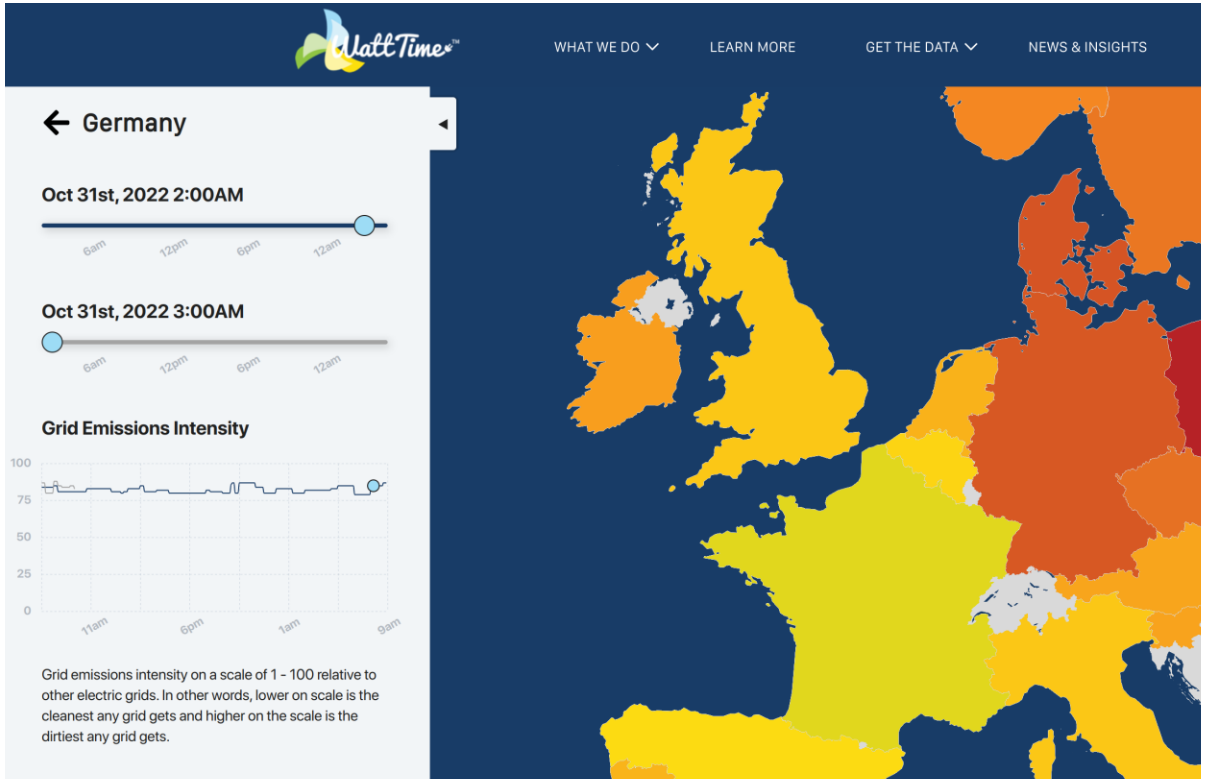
- Cloud regions vary significantly in terms of carbon emissions
- Google offers the **Region Picker** to take into account carbon footprint, price, and latency
- Region Picker does not take energy mix into account

  → Nuclear power plants are considered "low carbon“



Resource Selection – Locality

- **Watttime** as an alternative to Google Region Picker
- Watttime provides data on power plant emissions by using **measurements from space**
- Only useful if supplying power plant is known



Green Cloud Computing

- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in Cloud
- **Optimization measures for IaaS and PaaS**
  - Resource Selection – CPU & Location
  - **Scaling Strategies**
- Cloud Native Software Development
- Rebound Effects

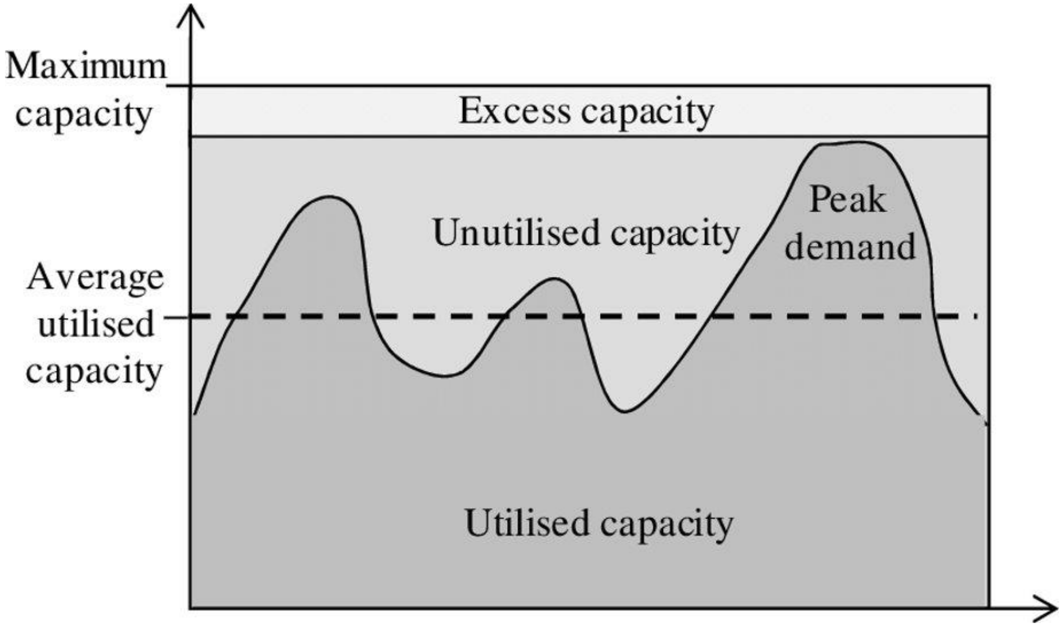
Overprovisioning

Anti-pattern when operating in the Cloud: **Overprovisioning**

- Permanent allocation of resources in order to be able to serve **peak loads**
- On average, resource allocation exceeds actual demand

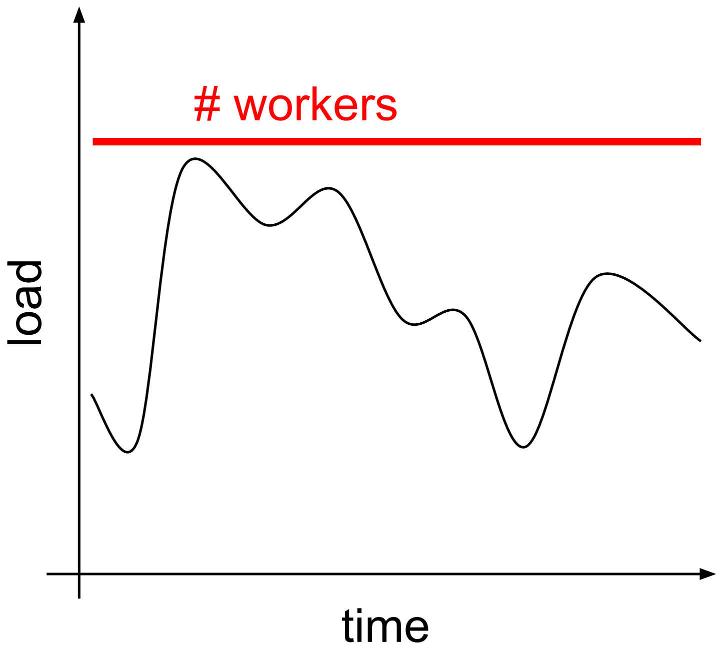
Examples:

- Provision online shop for peak loads in Christmas season
- Provision for execution of scheduled jobs



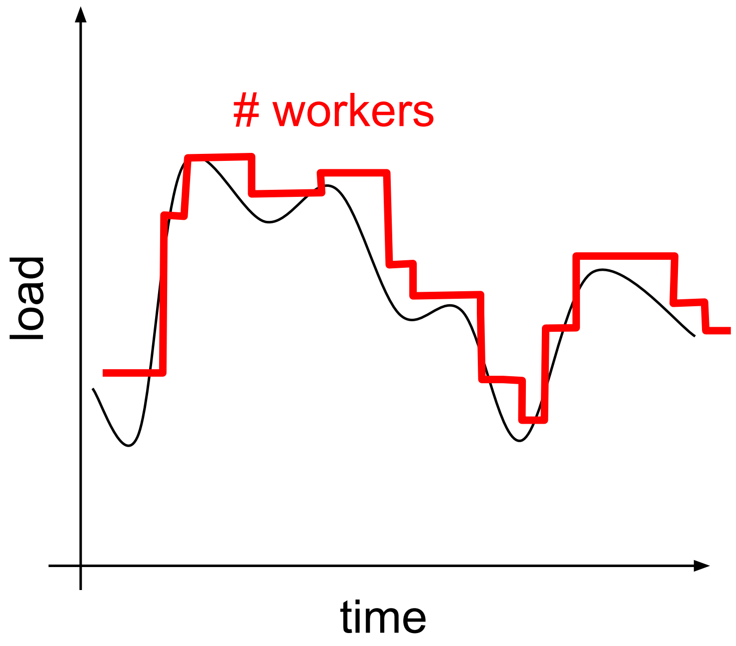
Overprovisioning

- Overprosioning often occurs when **static or reactive scaling** is used



Static Scaling

No elasticity → overprovisioning needed to handle peak loads



Reactive Scaling

Resources are not provided fast enough → overprovisioning needed

envite

Scaling Strategies

- Different scaling strategies to avoid overprovisioning and to save resources

Pro-Active / On-Prediction

Random / On-Coincidence

Demand Shifting / On-Availability

Demand Shaping / On-Availability

Scaling Strategies – Pro-Active

- Pro-active provisioning of resources
- Demand-driven scaling **before** actual demand is present ("On-Prediction")
- Counteracts overprovisioning that occurs due to excessive startup times of VMs or other instances

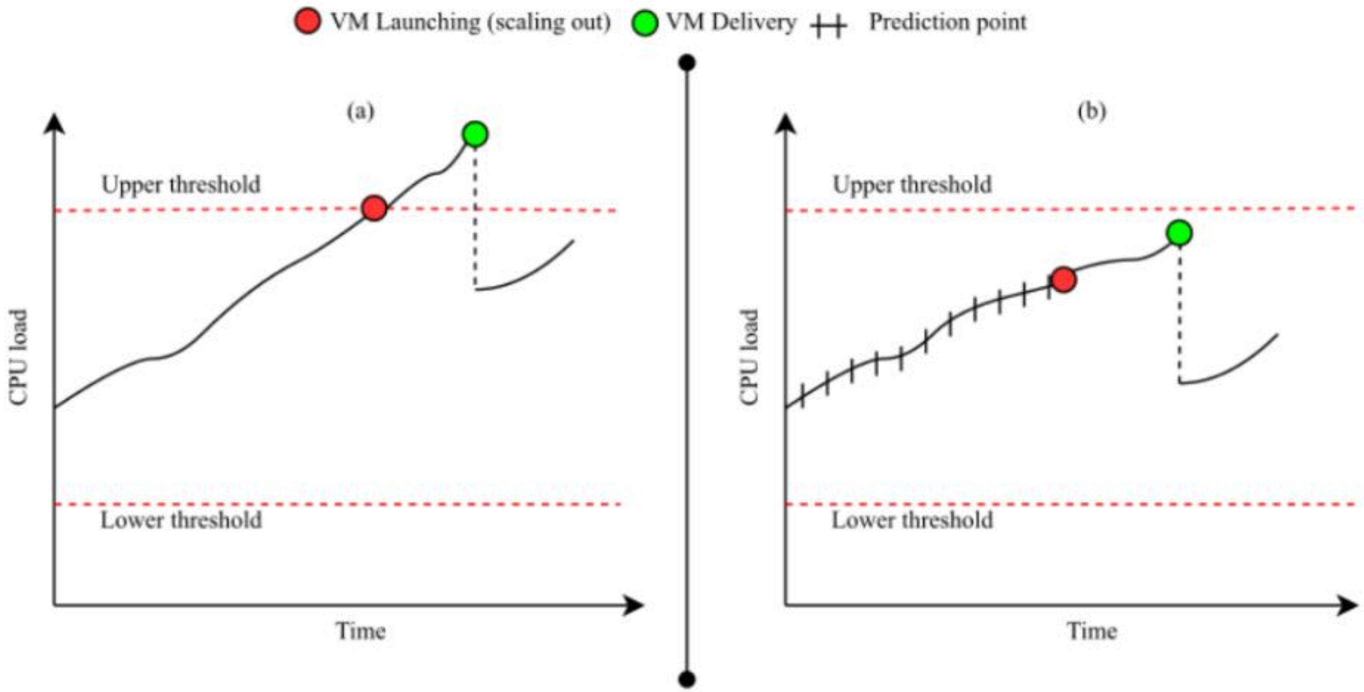


Fig. 1. Elasticity approaches: (a) reactive; (b) proactive.

Scaling Strategies – Random

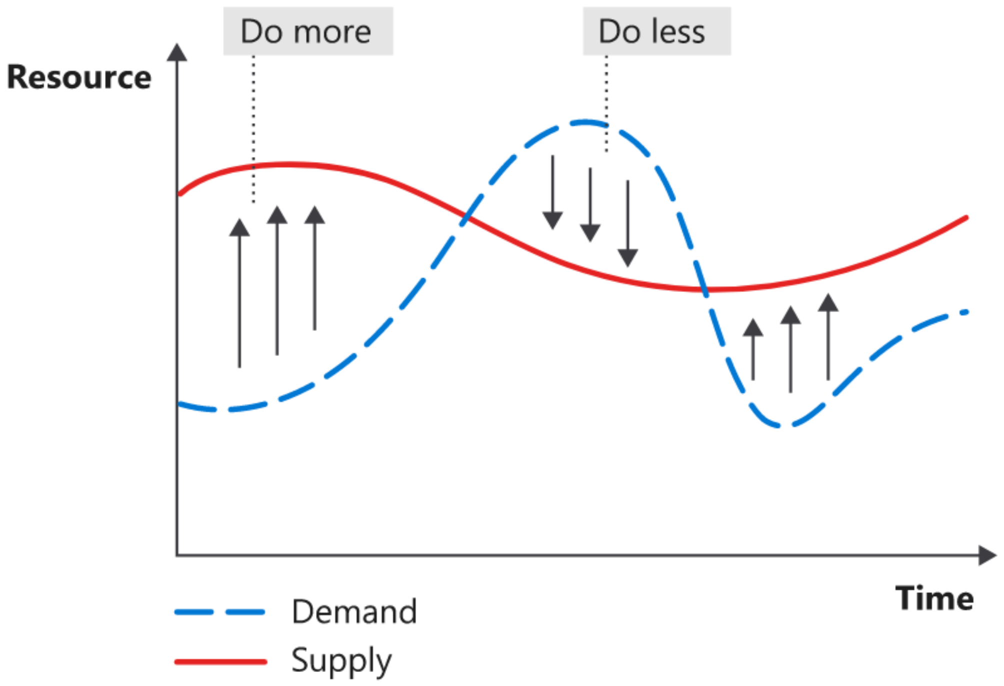
- **Random provisioning** of resources to equalize peak loads
- Regular workloads will be started at random times to better distribute the load on the system
- Only possible for workloads without user interaction

Examples:

- Event-Streaming applications
- Creating a database backup that would otherwise always run at 12 a.m.

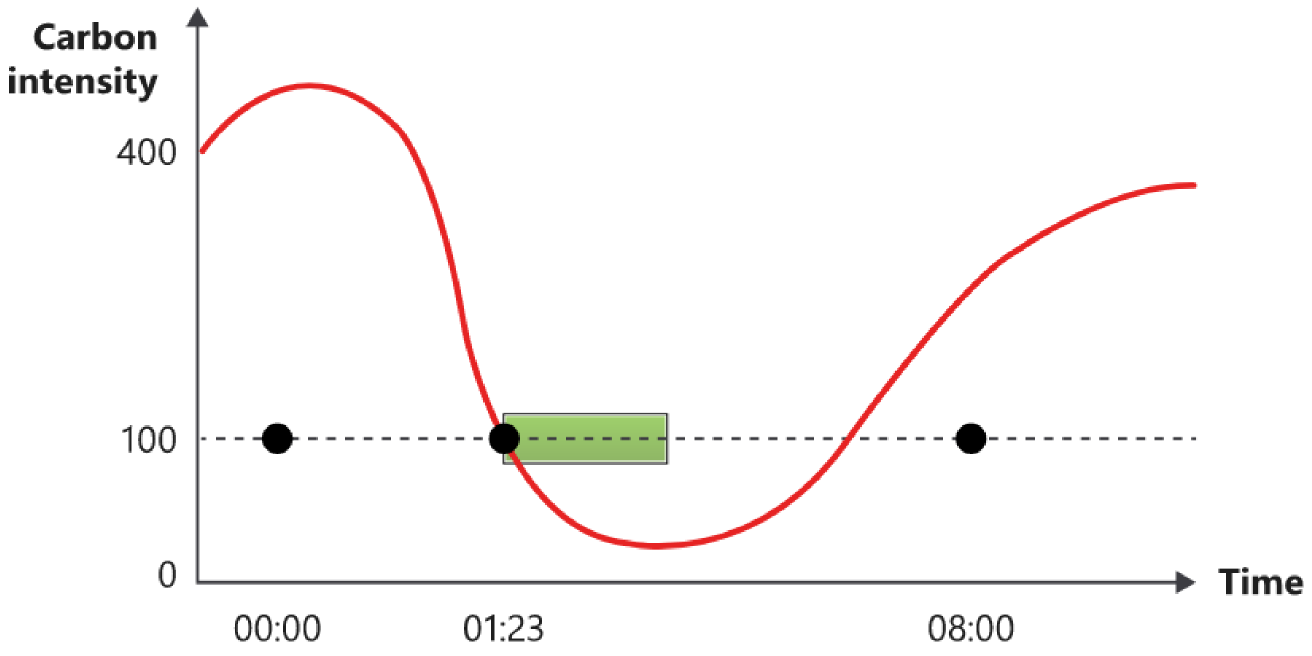
Scaling Strategies – Demand Shaping

- "On-availability" scaling shapes demand according to **available supply**
- Strategy for flexible workloads without time constraints
- Adjusts the provisioned resources to available resources
  - Examples for constrained resources: **CPU capacity, renewable energies …**
- Workloads can be matched to free capacities of the cloud provider with **on-demand, spot and preemptible instances**



Scaling Strategies – Demand Shifting

- Time-flexible workloads are shifted to **times or regions** where they can be executed with lower carbon emissions



- Alternatively, execution is shifted according to other criteria
  - Times or regions where unused cloud provider resources are available
  - Times or regions where own unused resources are available

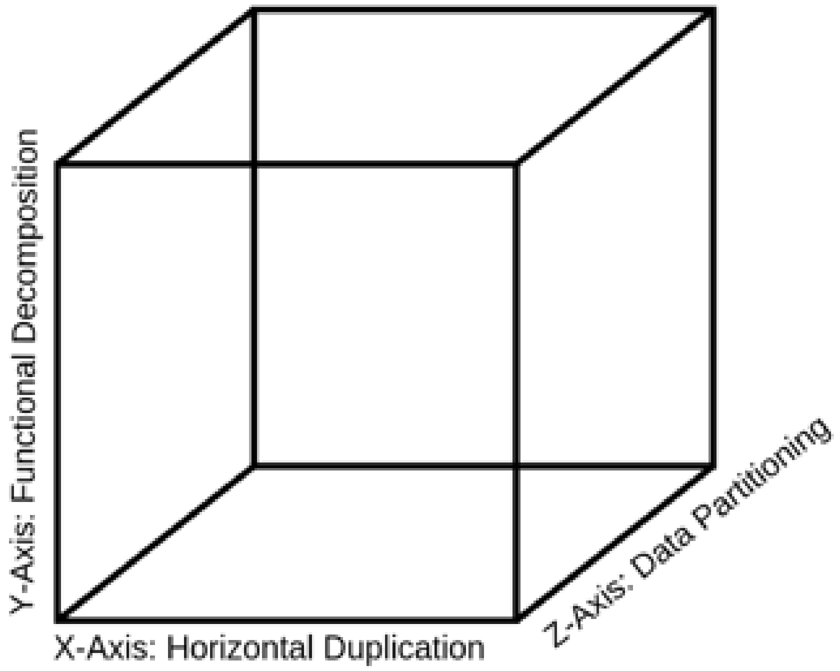
Green Cloud Computing

- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in the Cloud
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
- **Cloud Native Software Development**
- Rebound Effects

Cloud Native Software Development

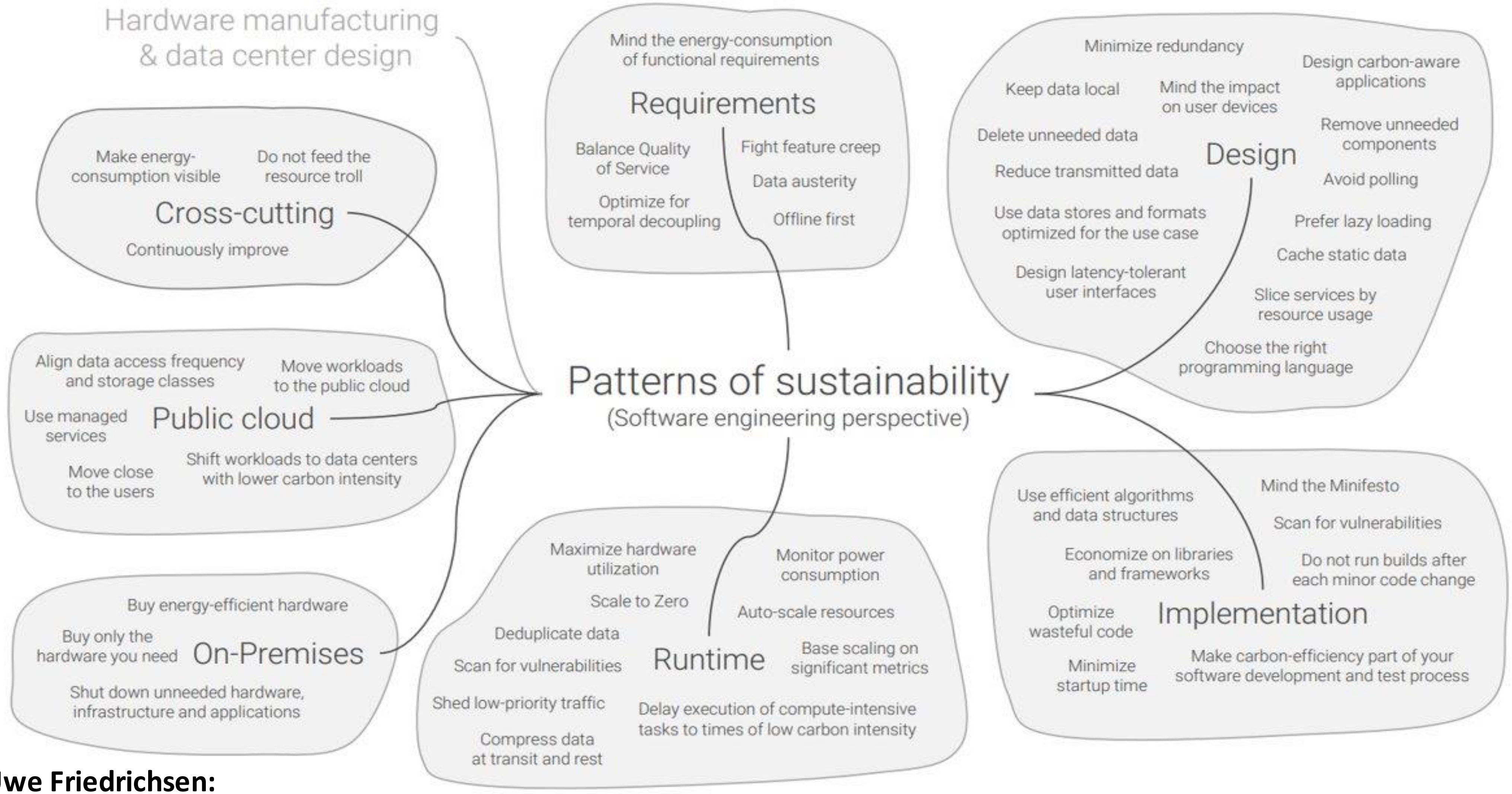
Prerequisites for taking advantage of all the benefits in terms of energy efficiency:

- **Fast startup times** for flexible scalability
- **Fast "graceful shutdowns"** to be able to shut down applications without data corruption
- **Available failover strategy** to get back online quickly
- Should be **stateless**
- Have good **scalability** according to the scale cube model



Applications should follow the **12 Factor Method** (https://12factor.net/)

Cloud Native Software Development



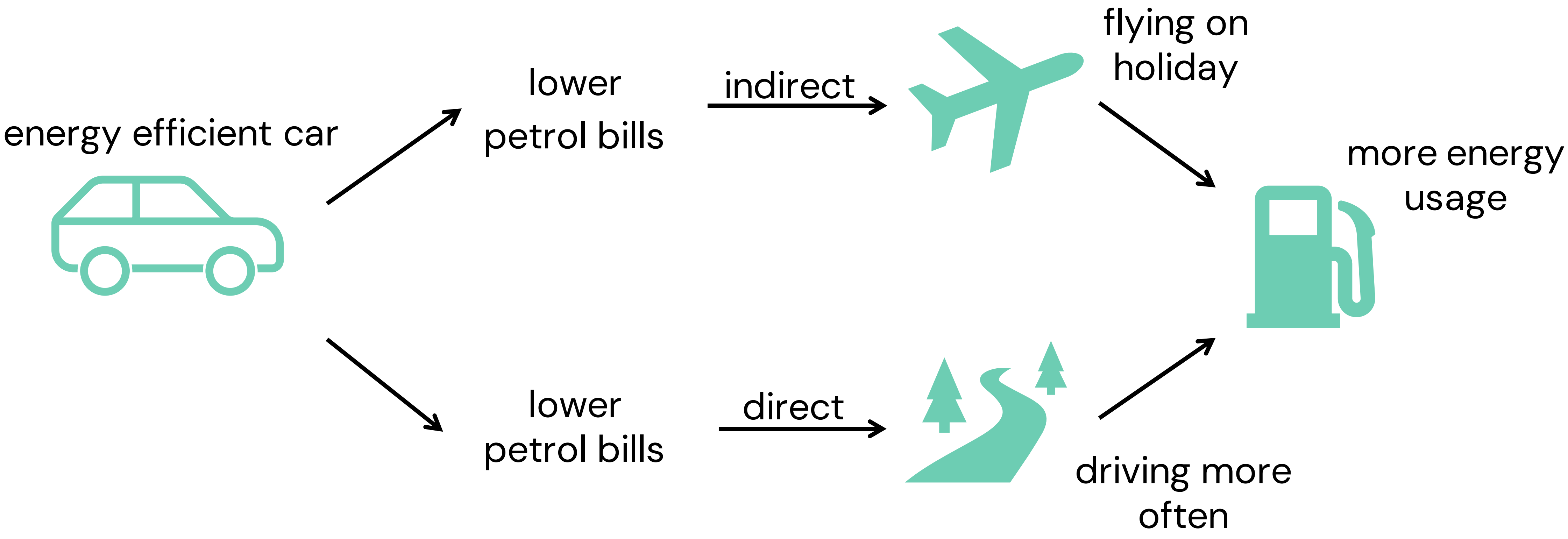
Talk by Uwe Friedrichsen:
https://speakerdeck.com/ufried/patterns-of-sustainability-going-green-in-it

Green Cloud Computing

- Energy Consumption by Data Centers
- Metrics & Sustainability of Cloud Providers
- Resource Utilization in the Cloud
- Optimization Measures for IaaS and PaaS
  - Resource Selection – CPU & Location
  - Scaling Strategies
  - Efficient Workload Distribution
- Cloud Native Software Development
- **Rebound Effects**

Rebound Effects

Rebound Effect = energy savings lead to changed behavior and increased energy usage

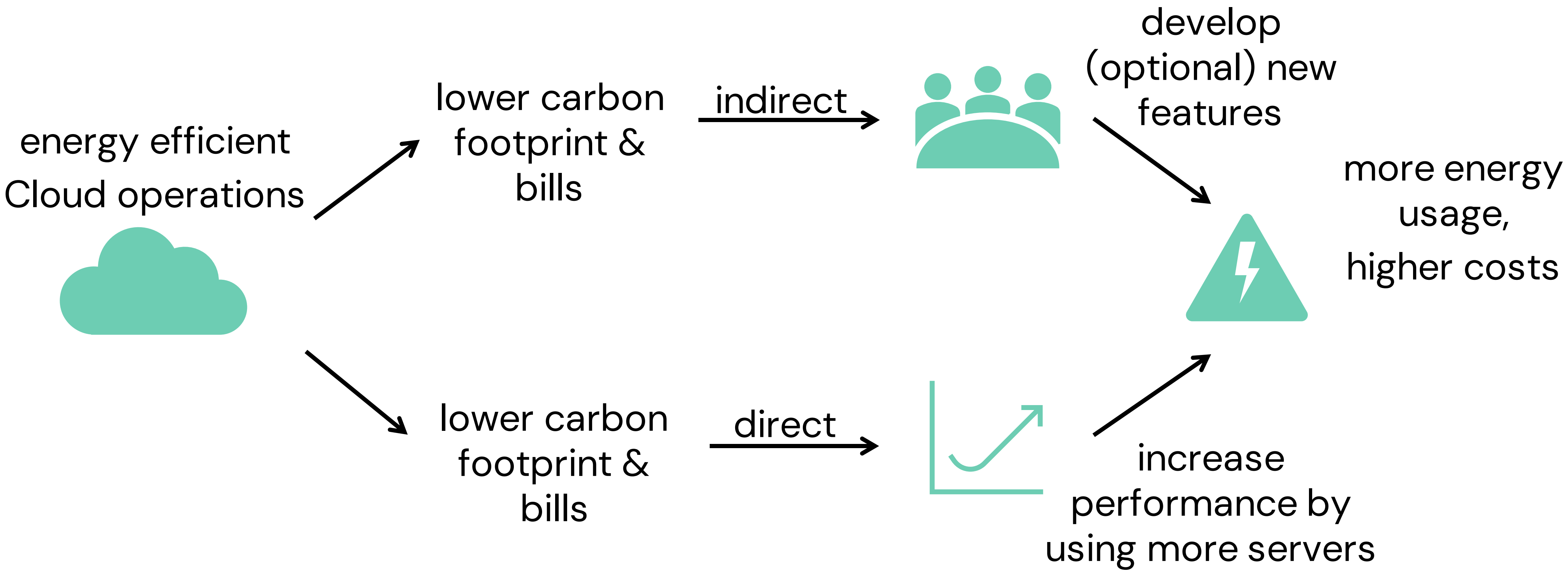


Adelmeyer, M.; Walterbusch, M.; Biermanski, P.; Seifert, K.; Teuteberg, F. (2017): Rebound Effects in Cloud Computing: Towards a Conceptual Framework, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 499-513

envite

Rebound Effects

- Cloud migration and optimizations lead to energy and costs savings
- **Risk**: savings encourage changed behavior and lead to increased energy usage



Green Cloud Computing

Can you think of any optimizations that could be made to your project from this semester to improve its energy efficiency?

Questions?

Thank you!

Uwe Eisele

uwe.eisele@envite.de

Nadja Hagen

nadja.hagen@envite.de

envite consulting GmbH

www.envite.de