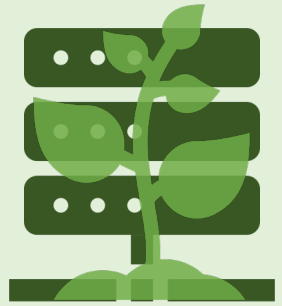
# Designing Cloud Servers for Lower Carbon



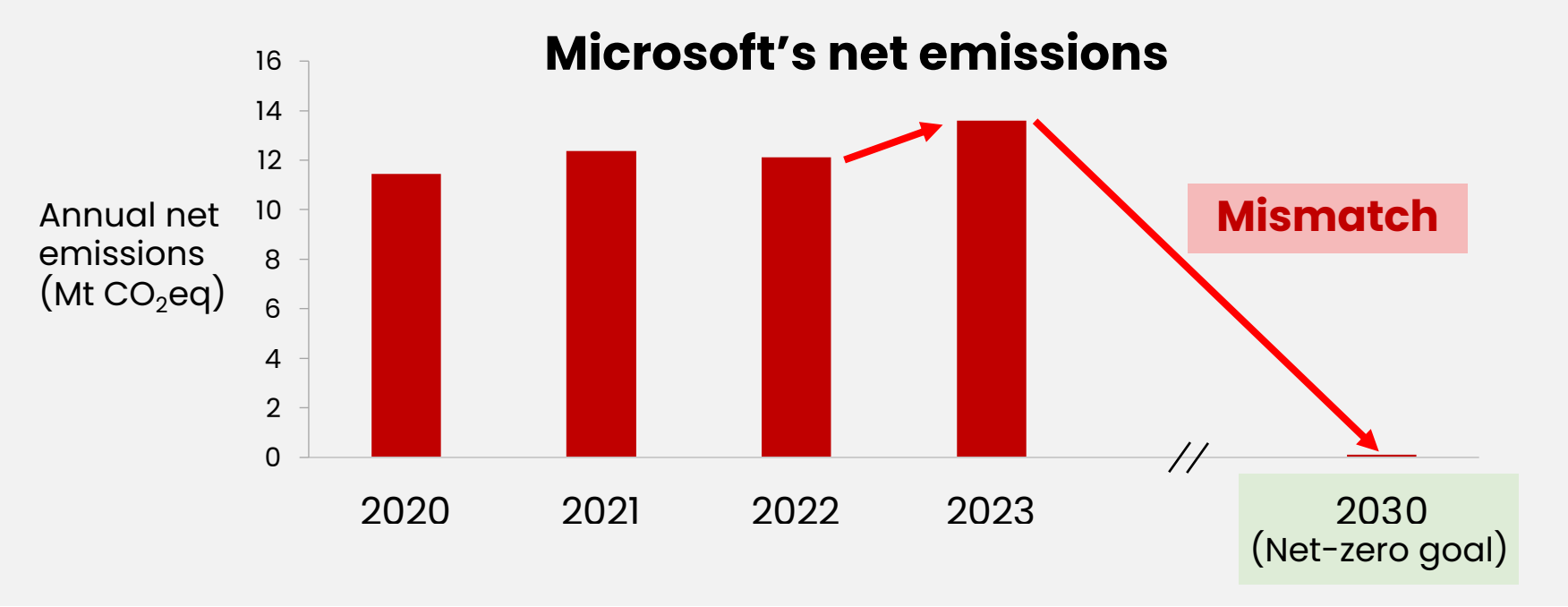
**Jaylen Wang**, Daniel S. Berger, Fiodar Kazhamiaka, Celine Irvene, Chaojie Zhang, Esha Choukse, Kali Frost, Rodrigo Fonseca, Brijesh Warrier, Chetan Bansal, Jonathan Stern, Ricardo Bianchini, Akshitha Sriraman

Carnegie Mellon University

Microsoft Azure

ISCA'24

# Cloud computing's carbon conundrum




*"The rise in our [...] emissions primarily comes from the construction of **more datacenters**..."*

[Microsoft Sustainability Report '24]

Cloud demand is growing, but carbon reductions are not keeping pace

# Where do emissions come from in data centers?

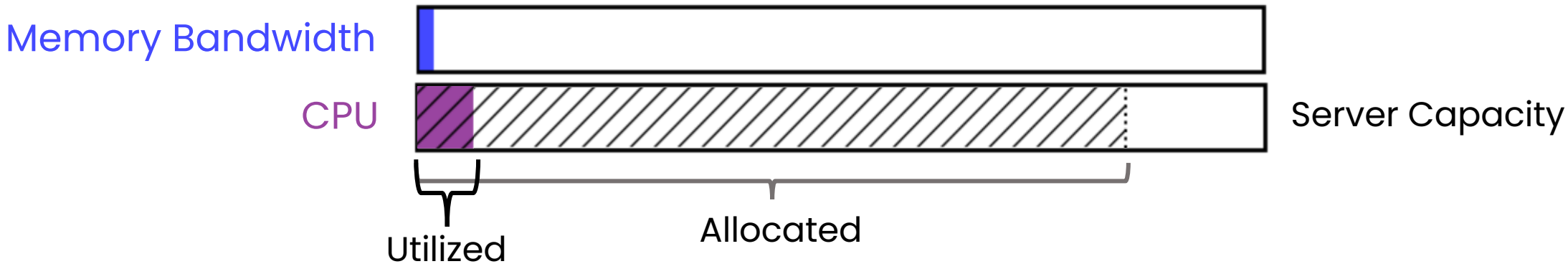
- Cloud computing contributes *two* types of emissions:


***Compute servers make up over 70% of operational and 40% of embodied emissions***

To reduce both types of emissions, we must lower the emissions of compute servers

# Why are compute servers carbon-inefficient today?

- Compute servers are underutilized:



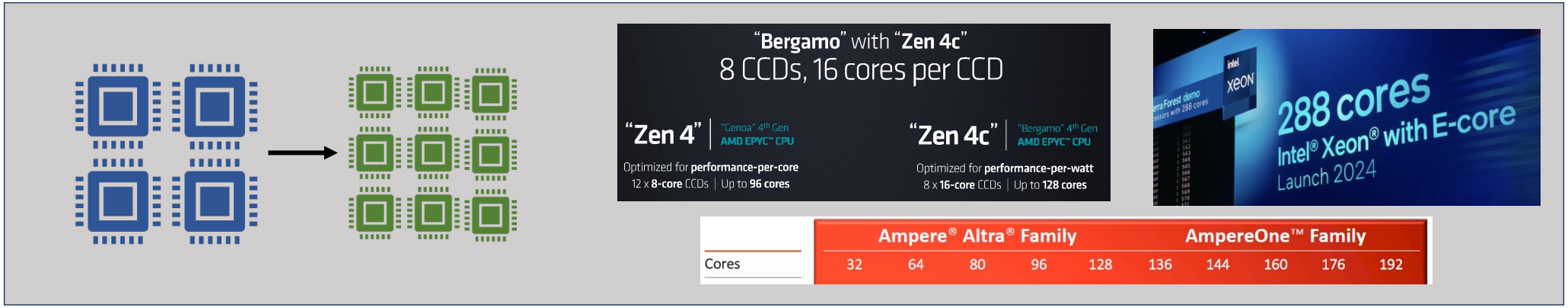

- Designing for higher performance loses out on potential carbon savings

There are opportunities to design lower carbon servers or ***GreenSKUs***

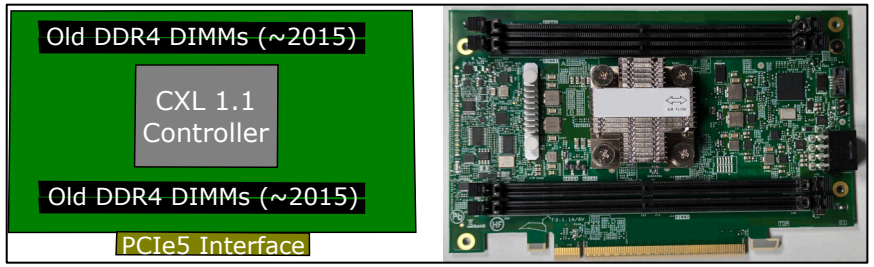
# What are the opportunities for designing **GreenSKUs**? Why now?

- Recent availability of deployable carbon optimizations:

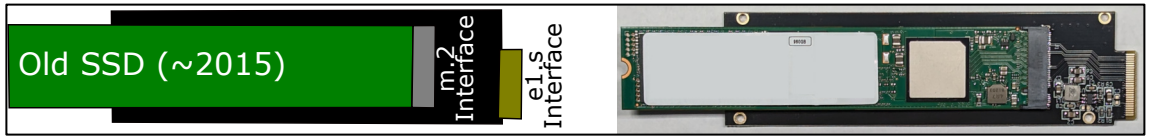
Efficient CPU cores




Memory reuse through Compute Express Link (CXL)



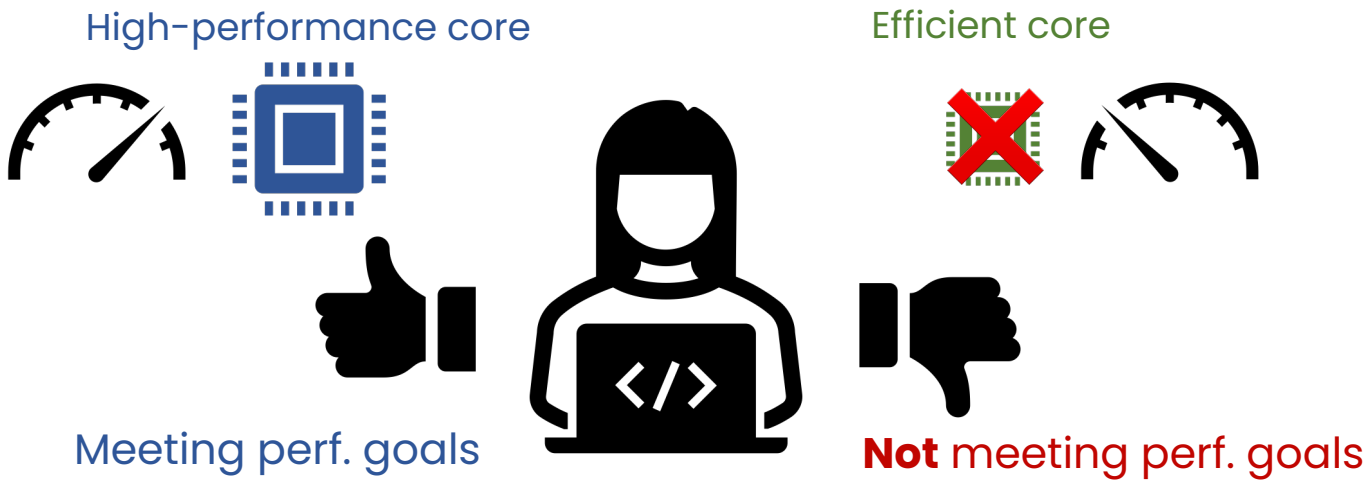

SSD reuse through PCIe backwards compatibility



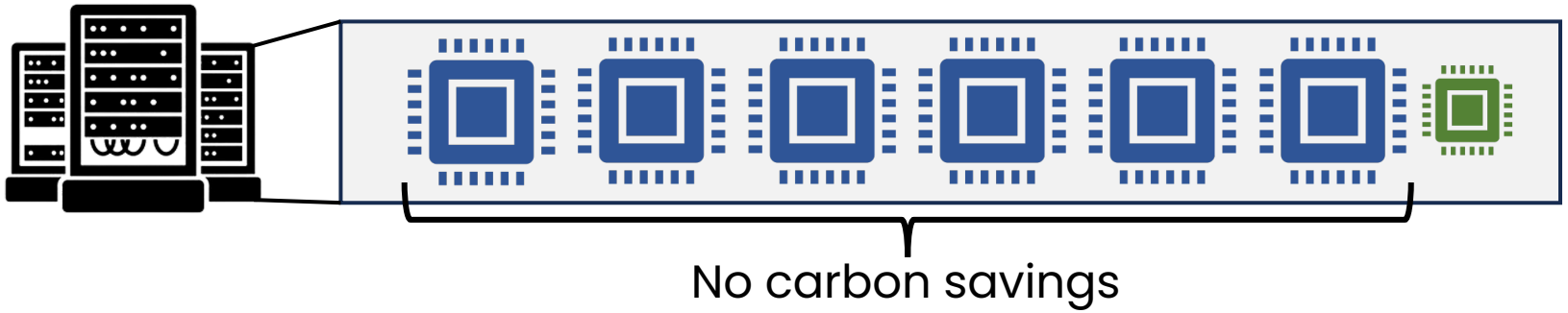

There's a recent availability of lower-carbon compute server components

# But, it's challenging to deploy **GreenSKUs** at scale…

- E.g., we may not meet performance goals due to carbon-efficient components:




Cloud deployment:




Once we design a **GreenSKU**, we need a **systematic** way to understand if it's worth it

# Contributions

1. A real **GreenSKU** **server design** and implementation

2. A **framework** to evaluate a **GreenSKU**'s carbon-saving potential

3. A carbon **evaluation** of our **GreenSKU** under Azure's production constraints

# Contributions

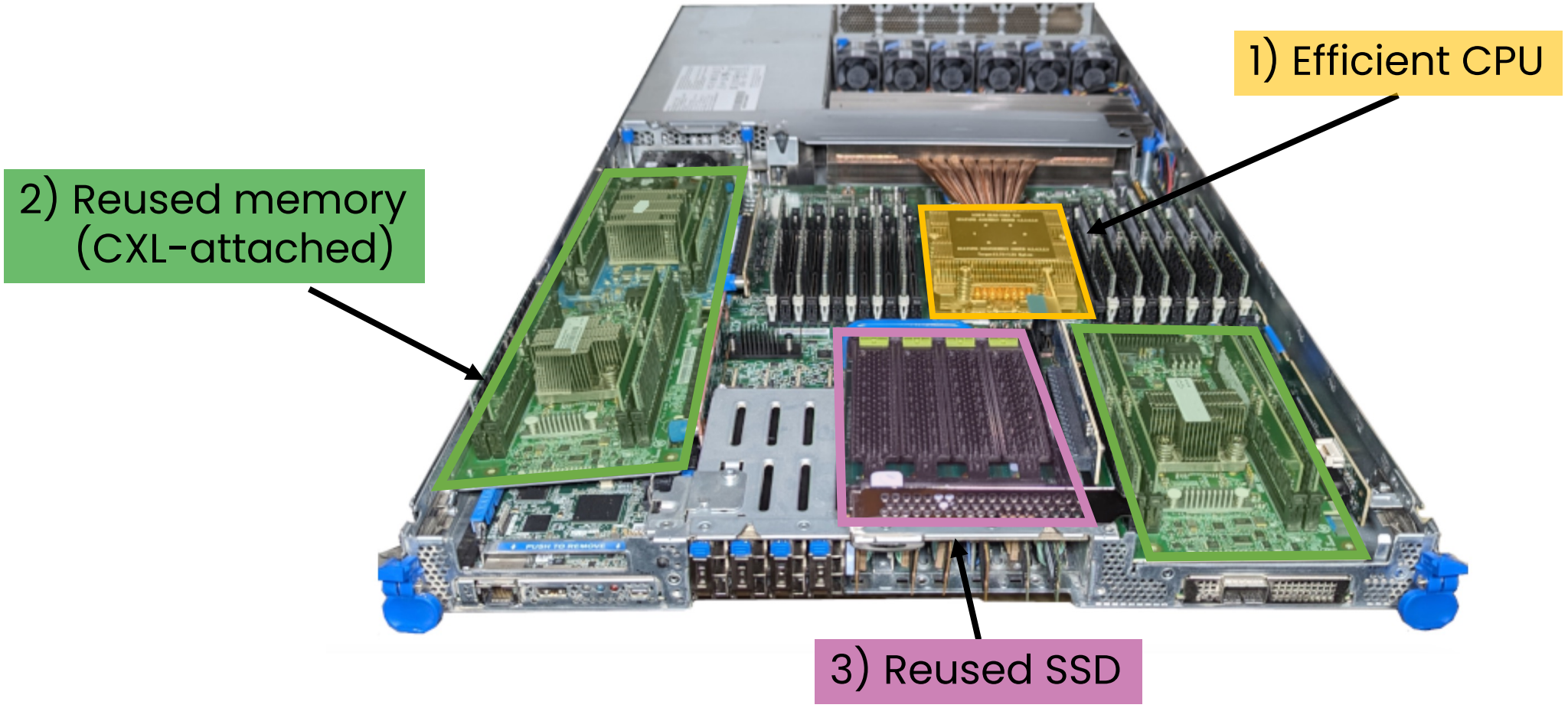
1. A real **GreenSKU** **server design** and implementation

2. A **framework** to taxonomize the considerations when evaluating a **GreenSKU**'s potential

3. A carbon **evaluation** of our **GreenSKU** with our framework

# Our **GreenSKU** server design

- We leverage recent carbon optimizations to design and build our **GreenSKU**



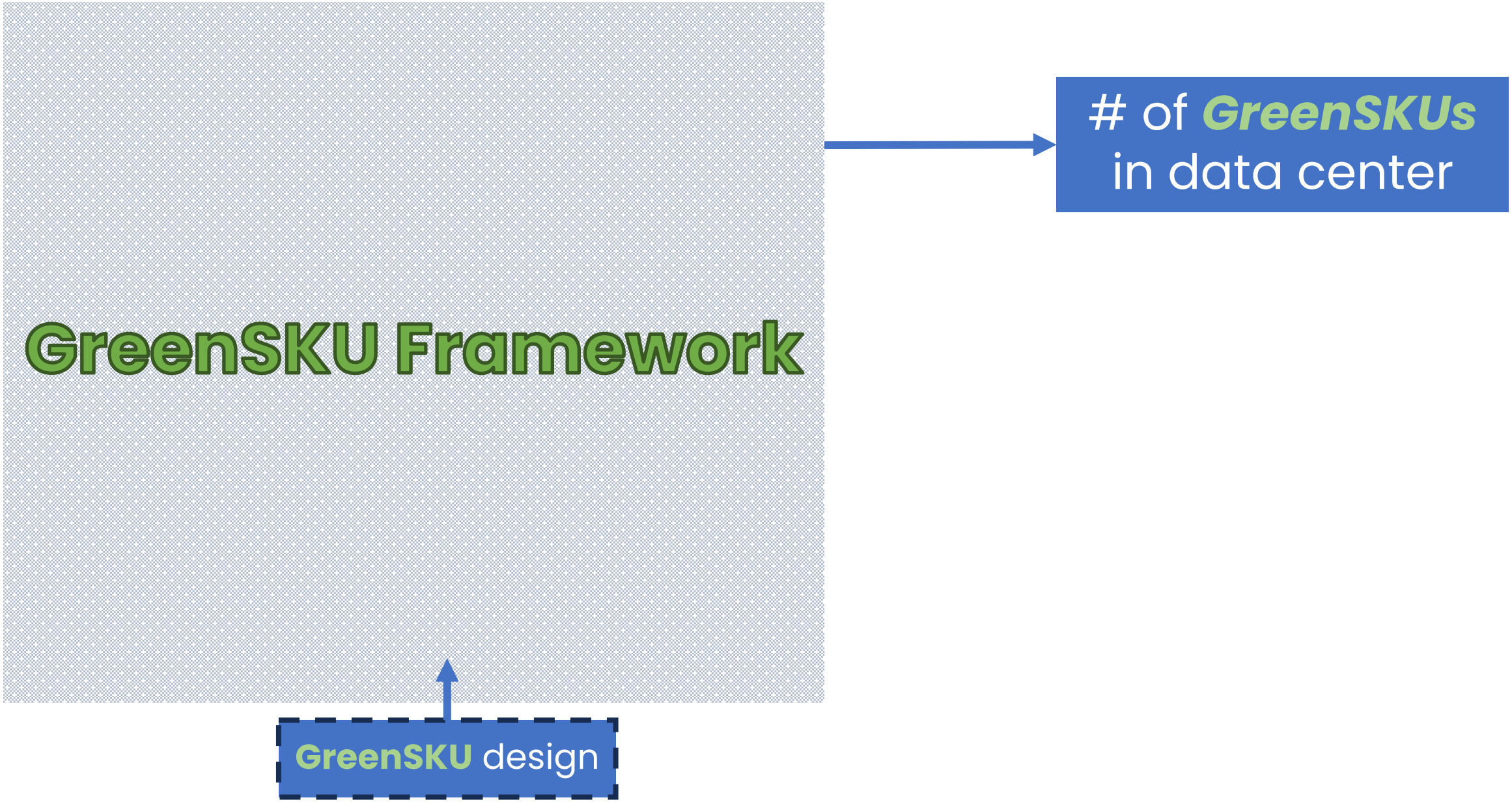
# Contributions

1. A real **GreenSKU** **server design** and implementation

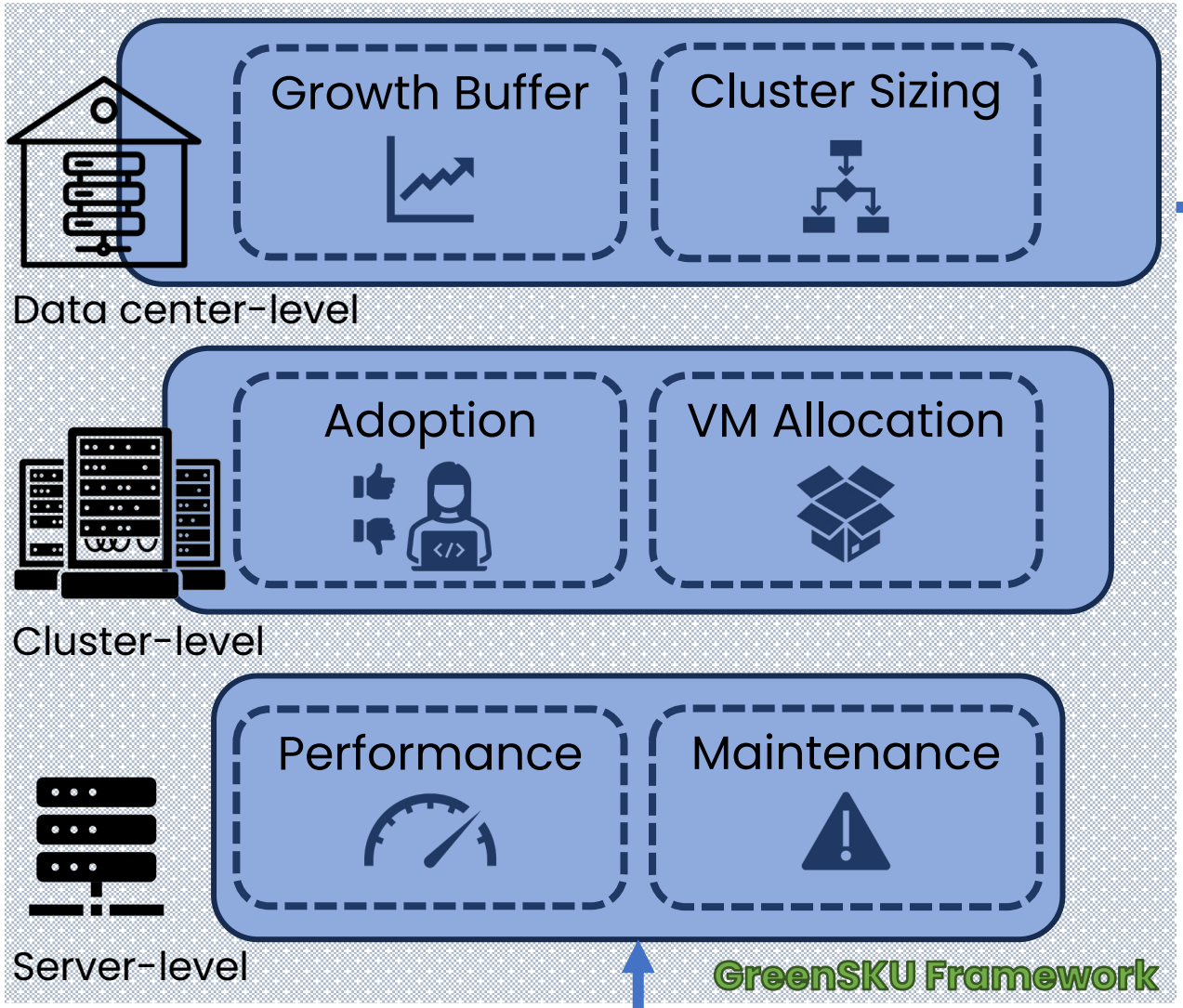
2. A **framework** to taxonomize the considerations when evaluating a **GreenSKU**'s potential

3. A carbon **evaluation** of our **GreenSKU** with our framework

# What is the framework's high-level goal?



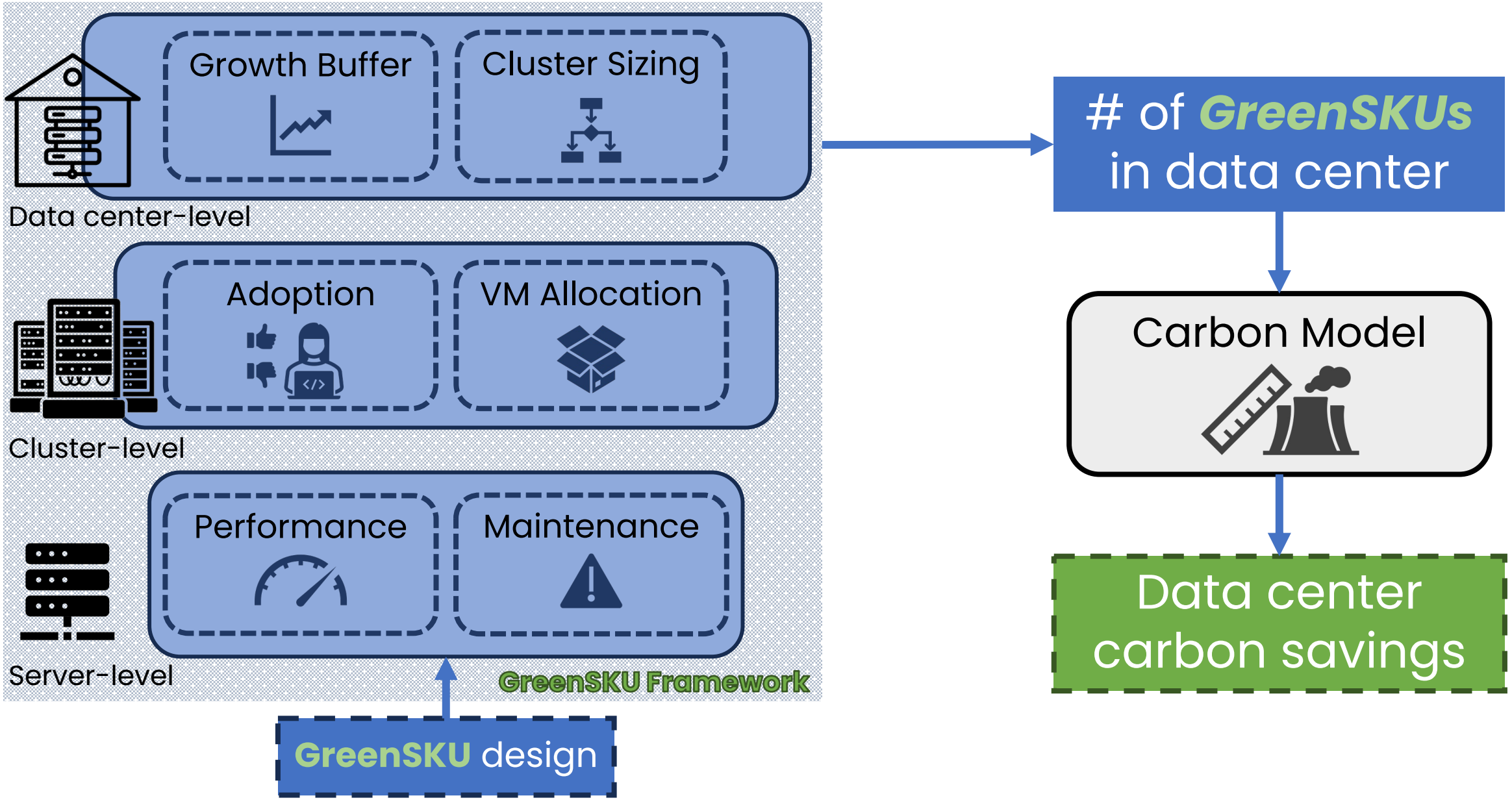
# What is the framework's high-level goal?



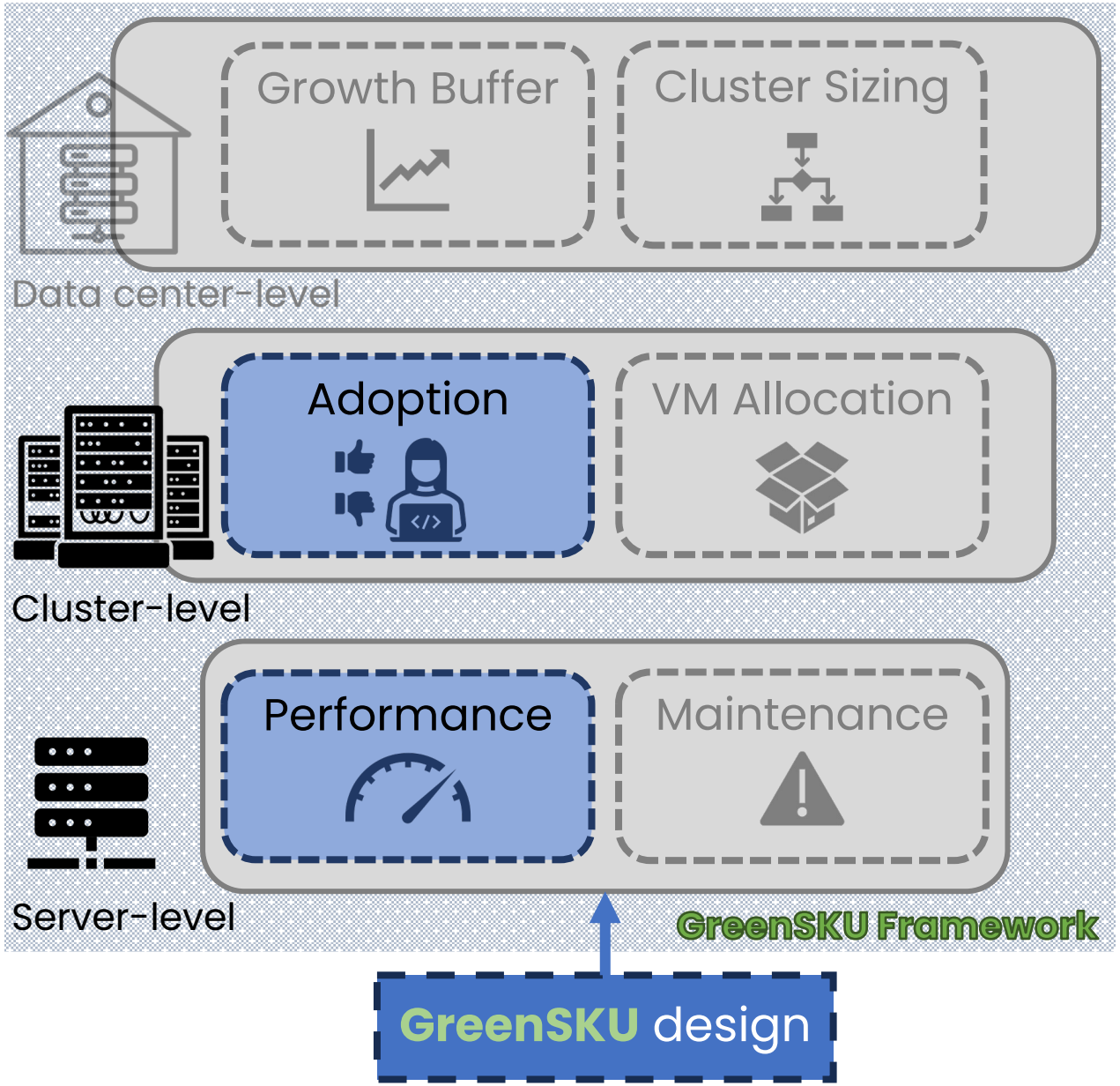

# of ***GreenSKUs*** in data center

**GreenSKU** design

# How much carbon do we save from these **GreenSKUs**?

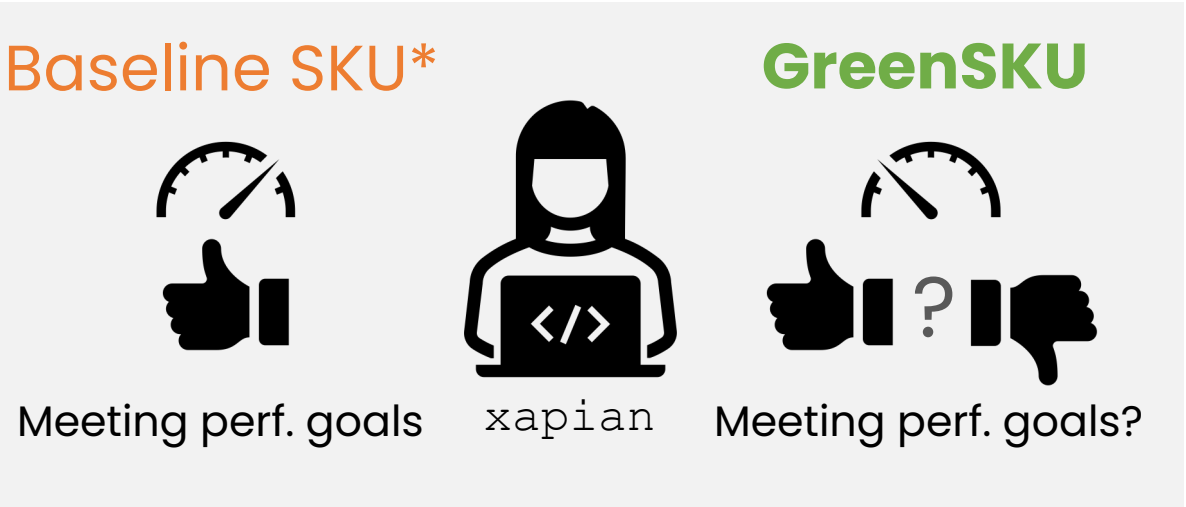


# How does a **GreenSKU**'s **performance** impact its **adoption**?



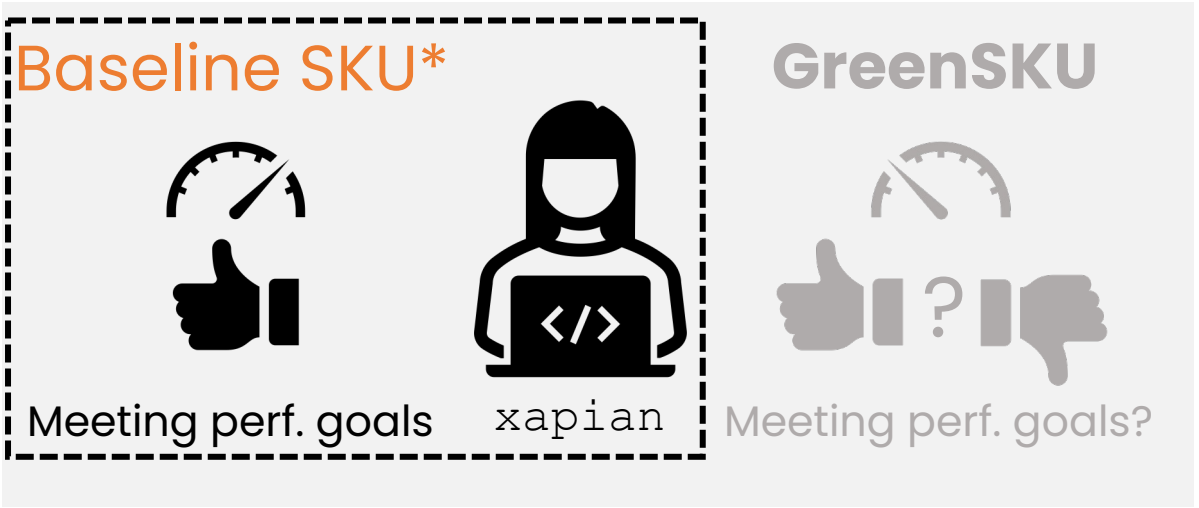
# How do we measure a **GreenSKU**'s **performance**?


**a high-performance deployed SKU without any **GreenSKU** optimizations*

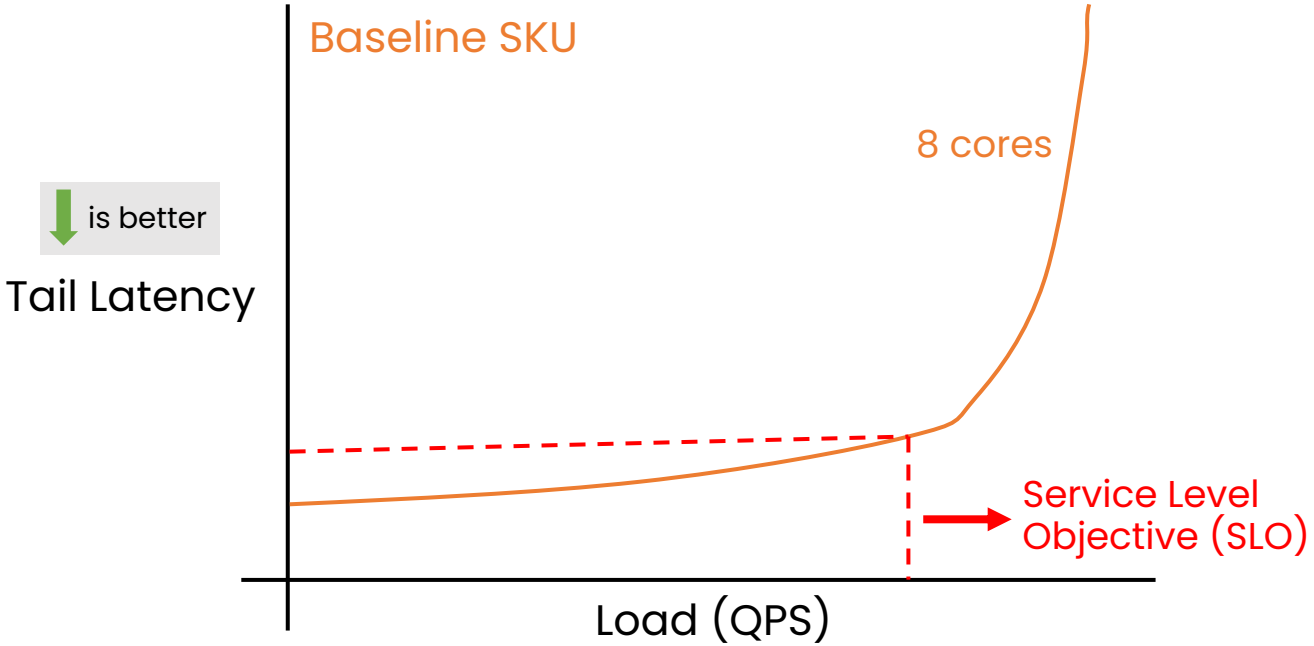


# How do we measure a **GreenSKU**'s **performance**?


**a high-performance deployed SKU without any **GreenSKU** optimizations*

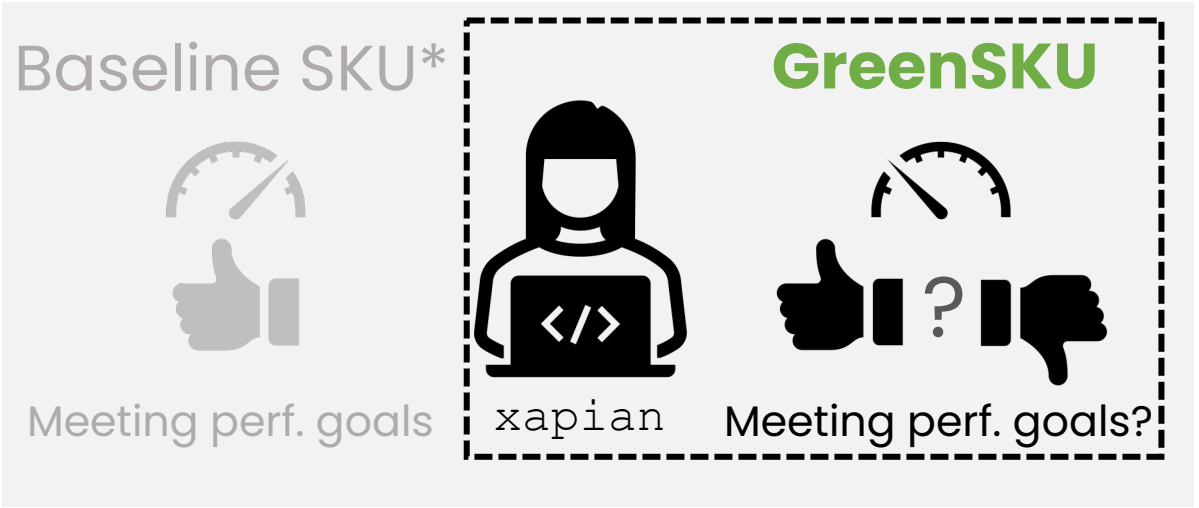



`xapian` Performance:

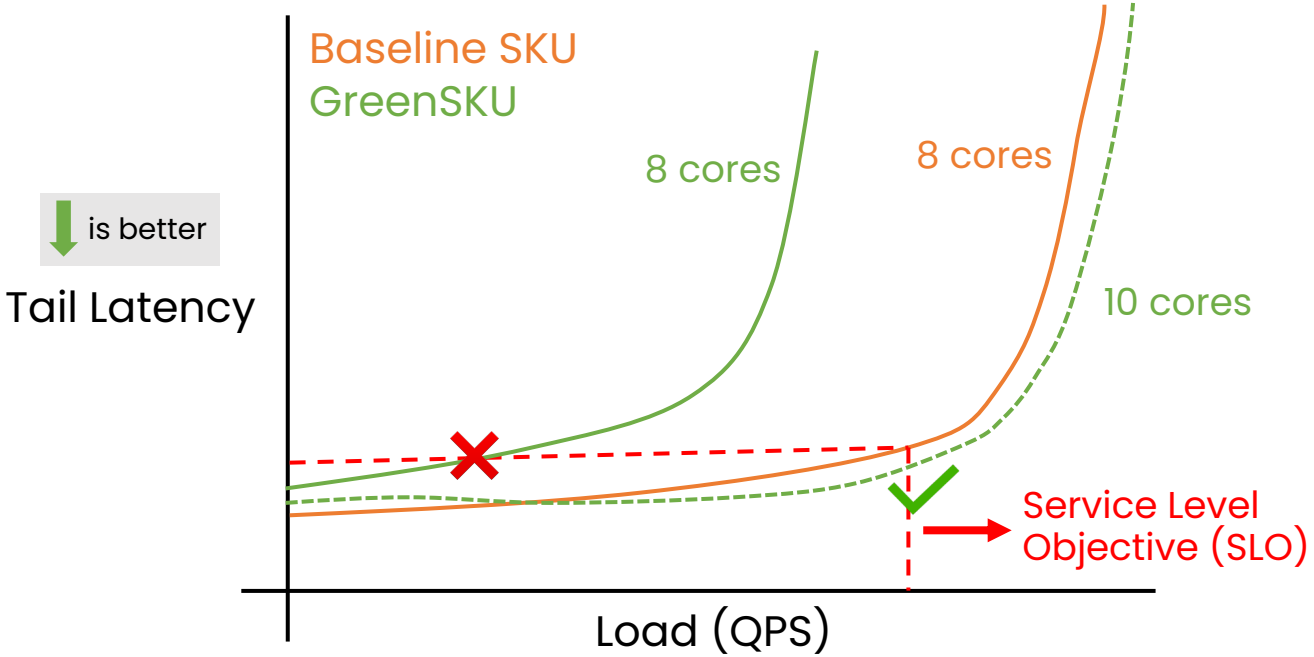


# How do we measure a **GreenSKU**'s **performance**?


*a high-performance deployed SKU without any **GreenSKU** optimizations*




`xapian` Performance:



# How do we measure a **GreenSKU**'s **performance**?


*a high-performance deployed SKU without any **GreenSKU** optimizations*

**Baseline SKU***

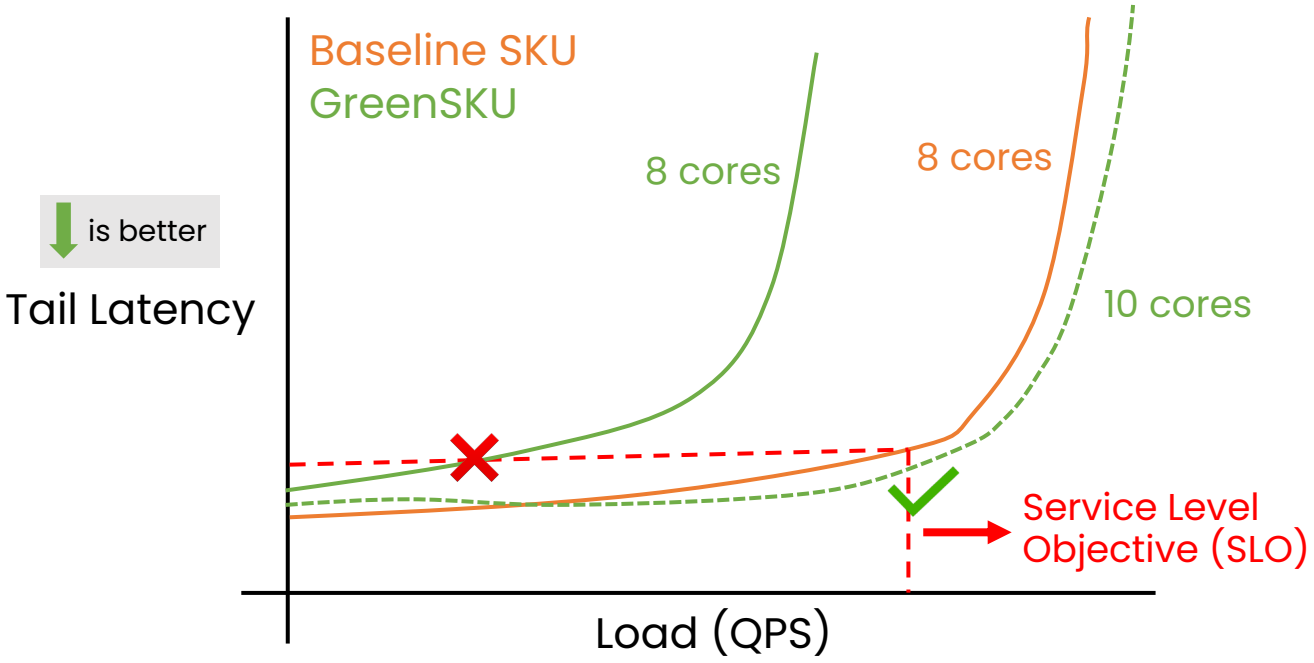
Meeting perf. goals

`xapian`

**GreenSKU**
**+25% cores**

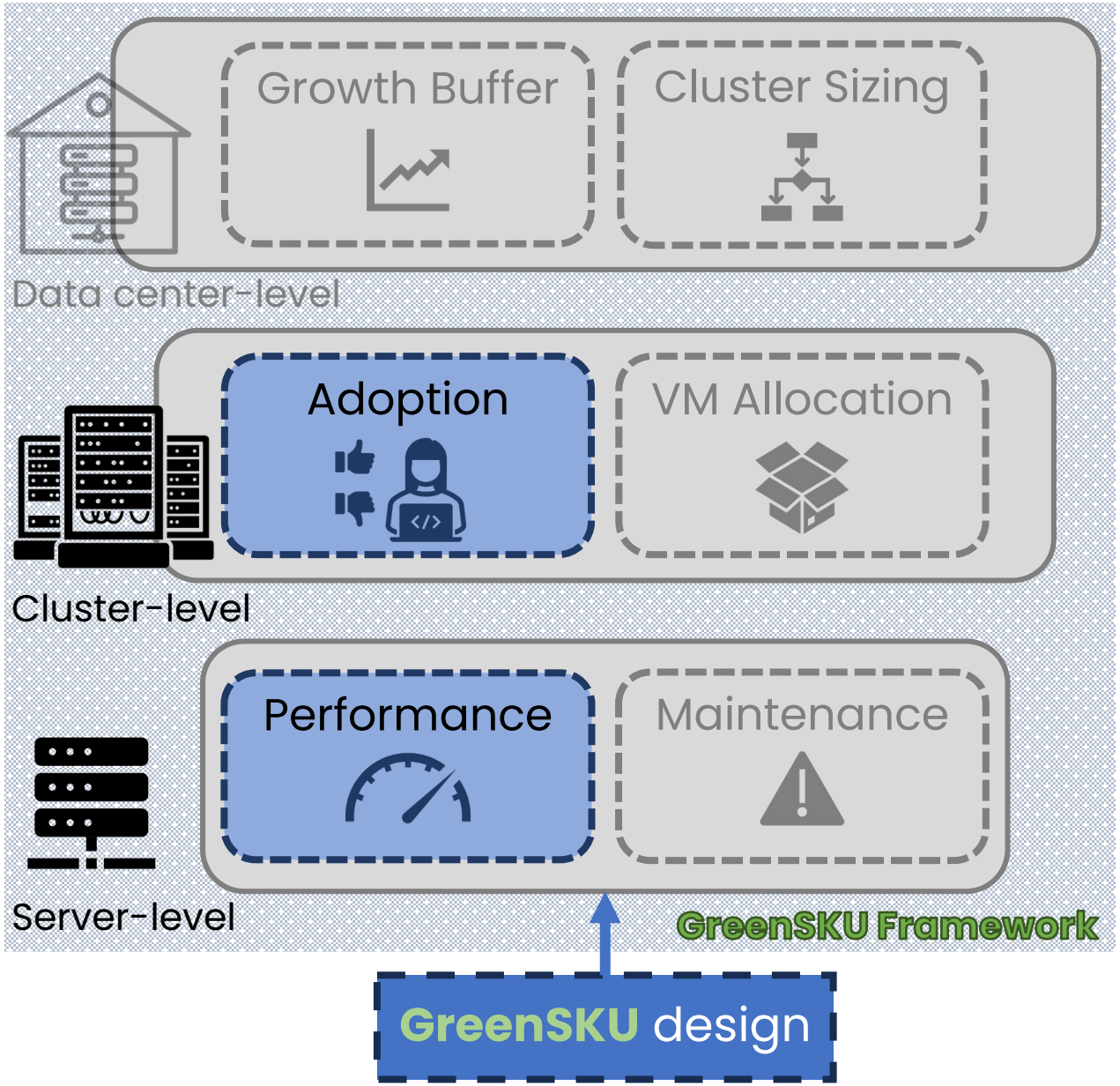
Meets perf. goals

`xapian` Performance:

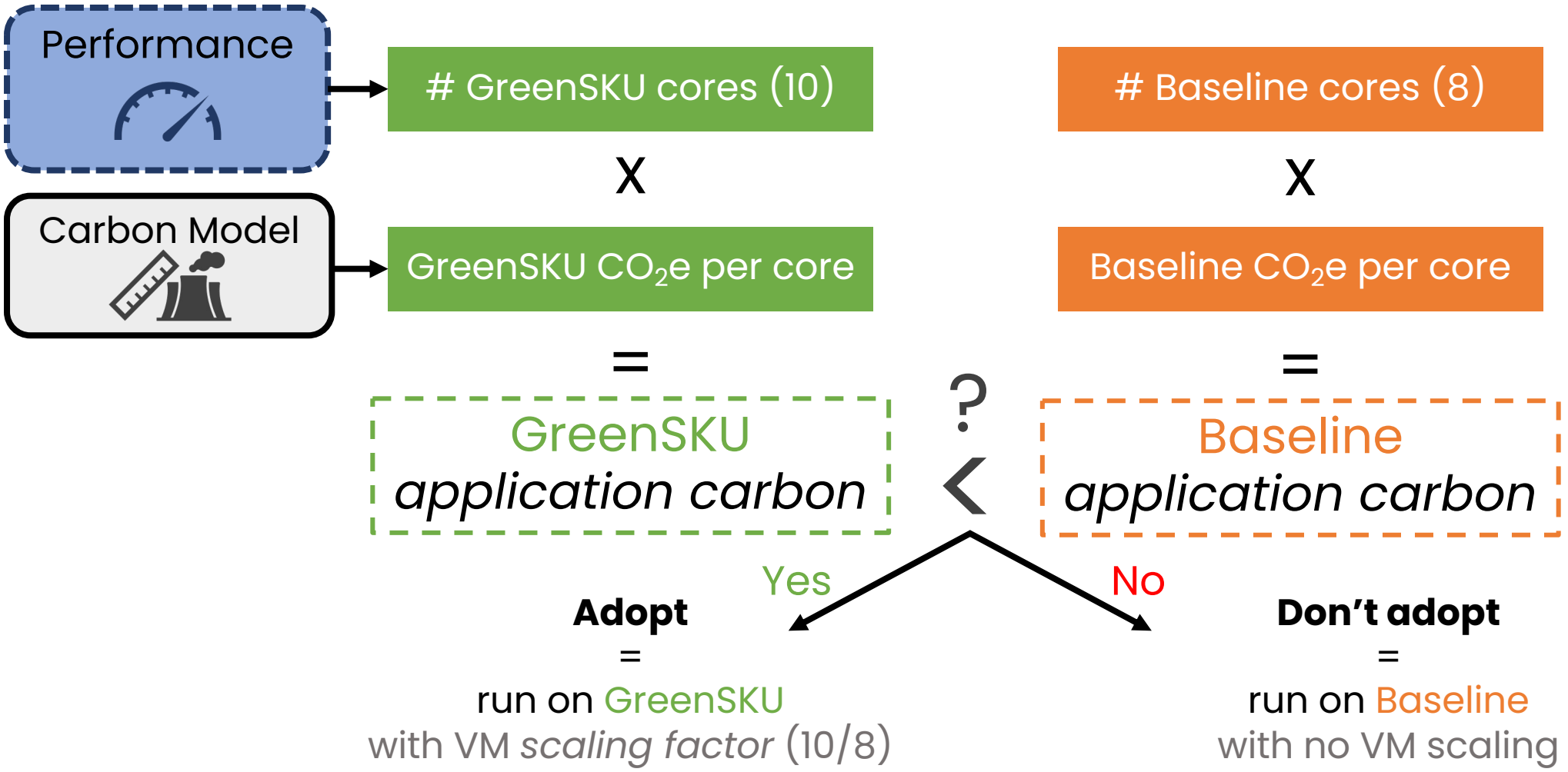


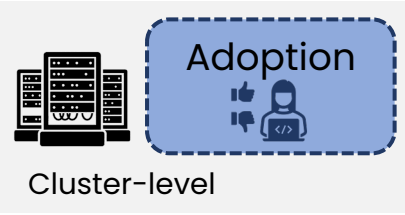

We identify the # of **GreenSKU** cores needed to achieve the baseline SKU's performance

# When should an application **adopt** a **GreenSKU**?



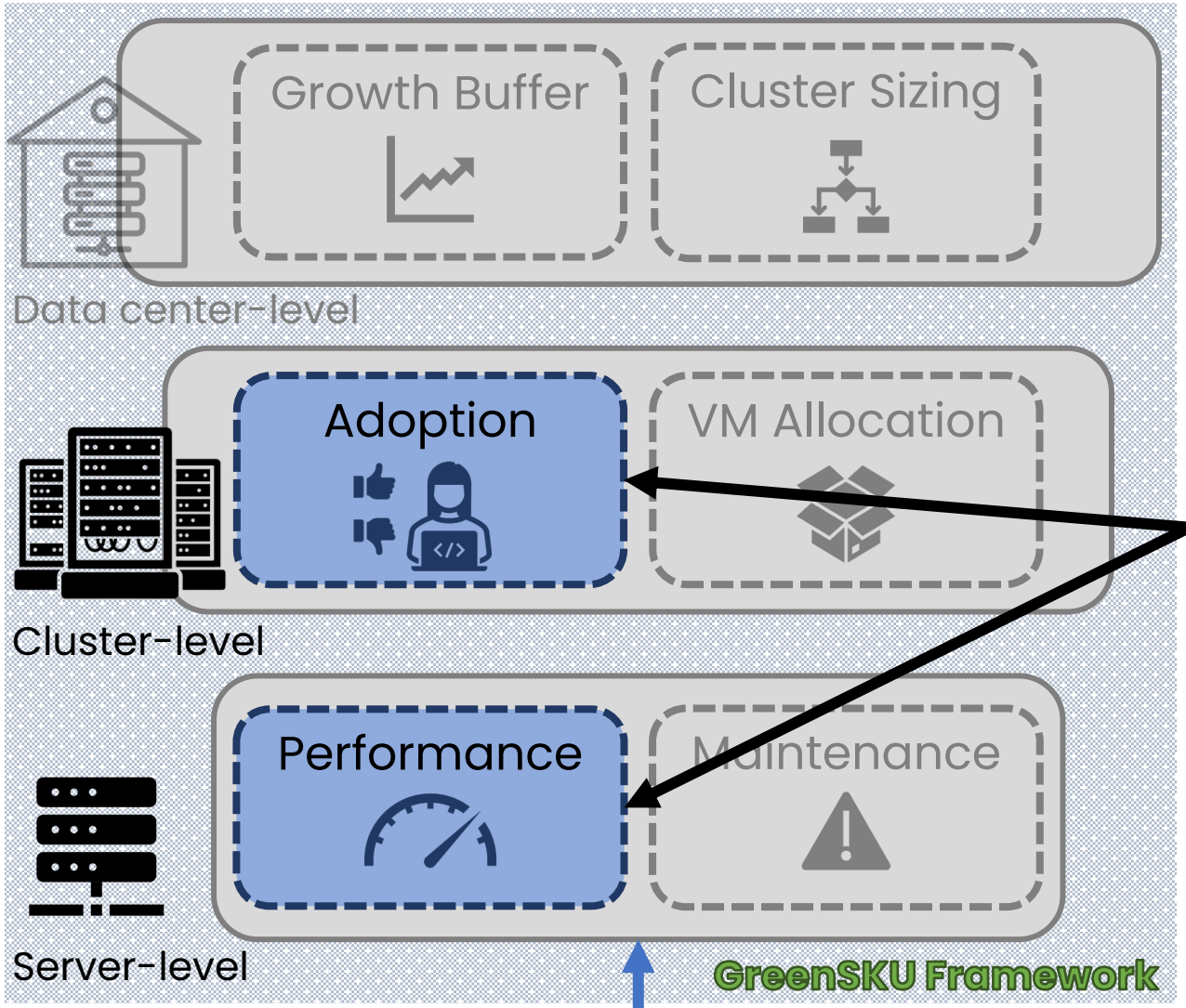
# When should an application **adopt** a **GreenSKU**?

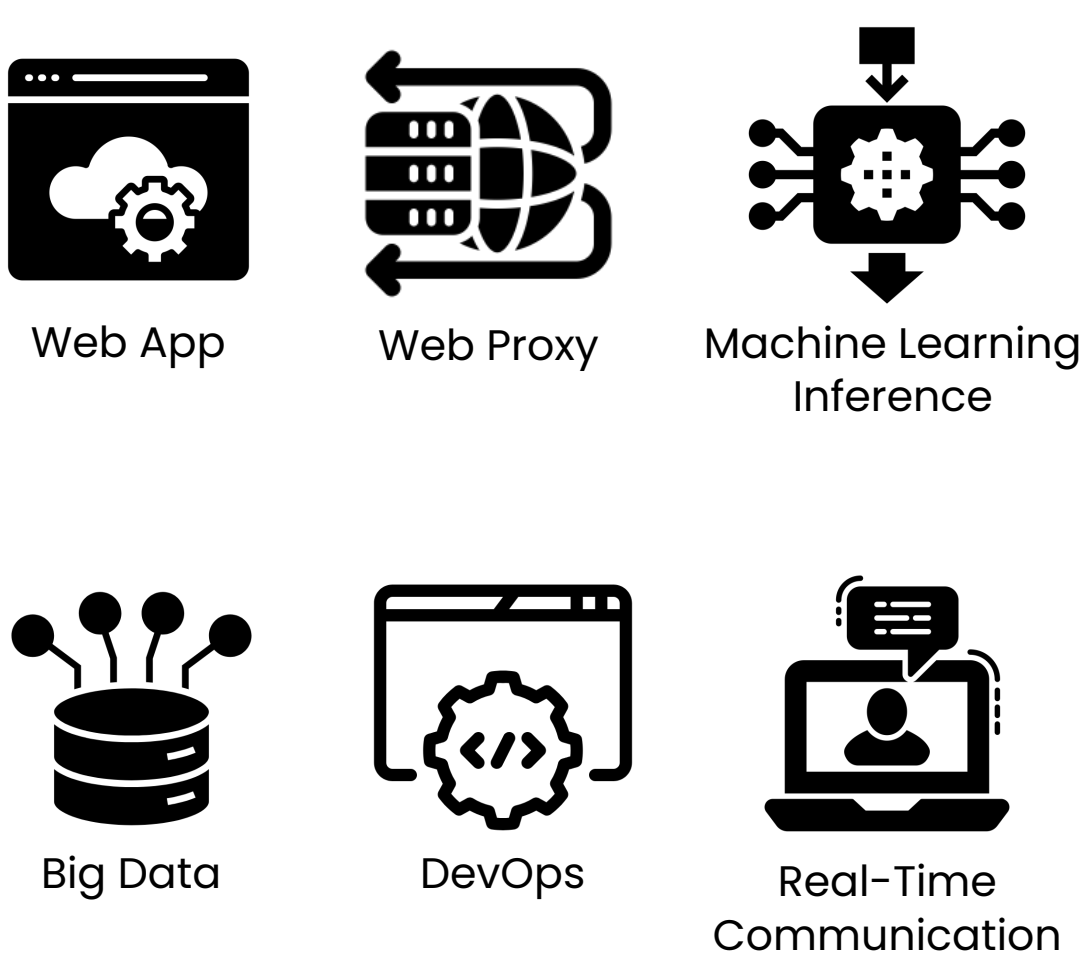


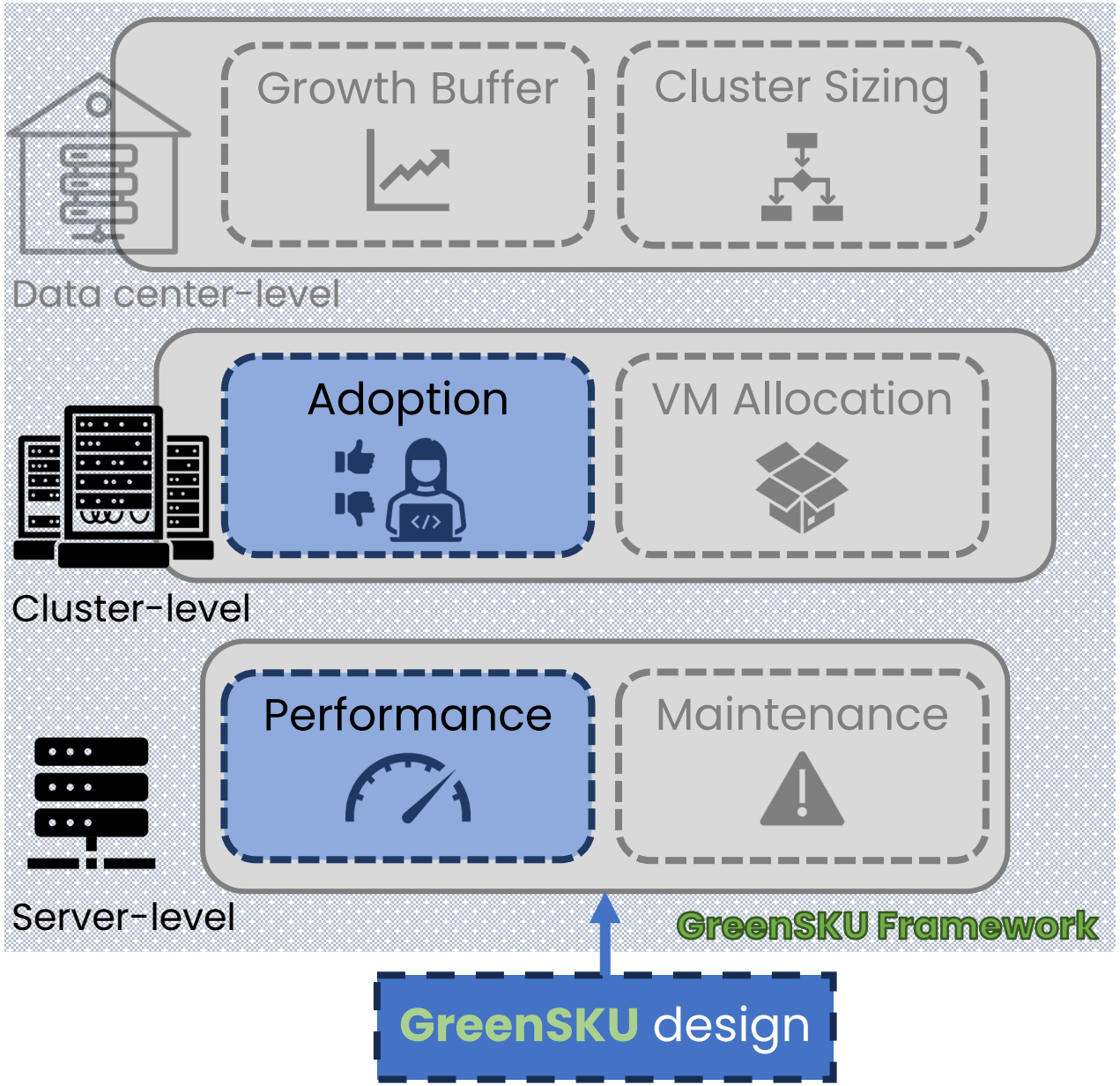

We run an application on the **GreenSKU** if it meets performance goals *and* saves carbon

# We study a **GreenSKU**'s potential across Azure's applications

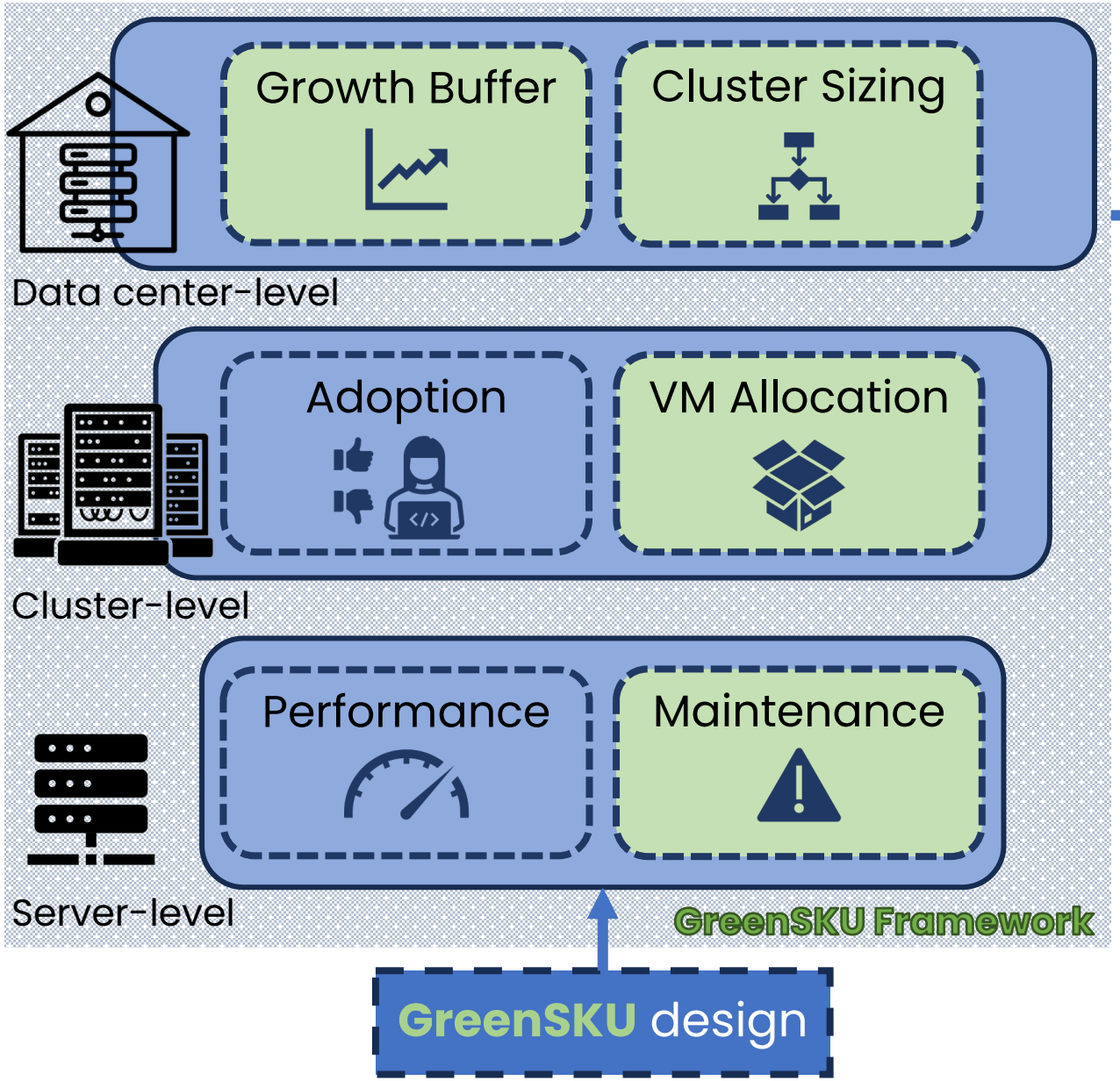


# What else impacts a **GreenSKU**'s carbon saving potential?



# What else impacts a **GreenSKU**'s carbon saving potential?




Data center carbon savings

See the paper!

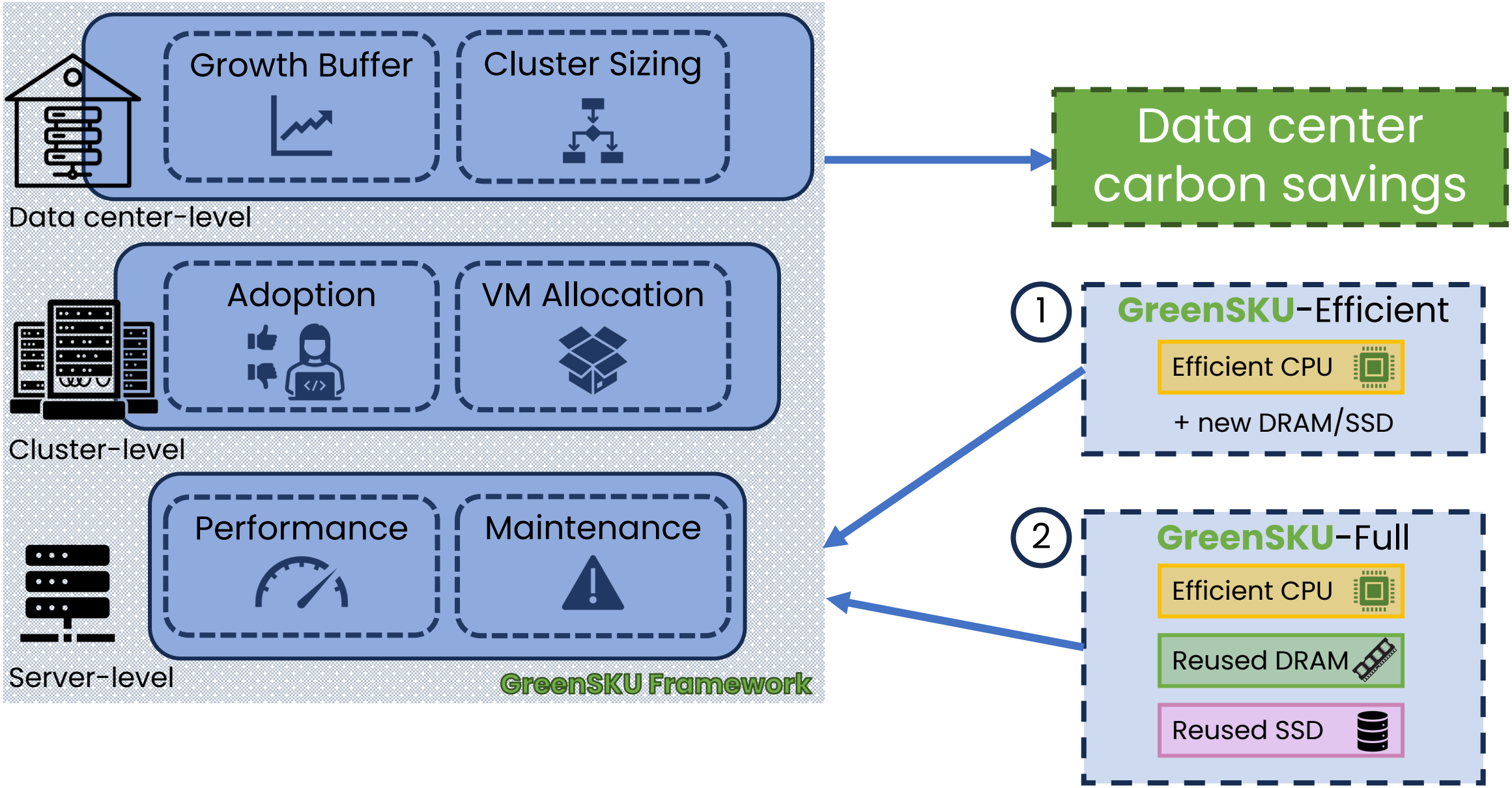
# Contributions

1. A real **GreenSKU** **server design** and implementation

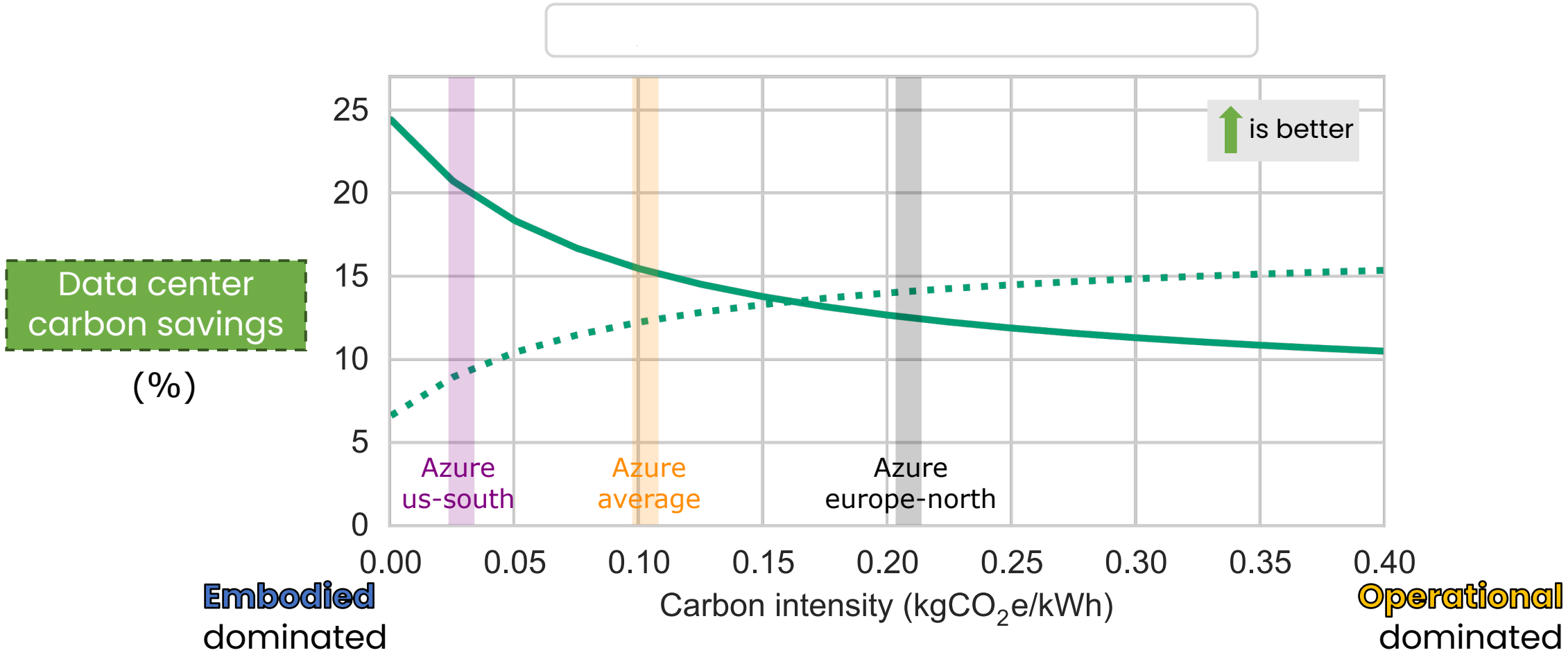
2. A **framework** to taxonomize the considerations when evaluating a **GreenSKU**'s potential

3. A carbon **evaluation** of our **GreenSKU** with our framework

# Evaluating two **GreenSKU** designs…



# How much do our **GreenSKUs** save data center-wide?




(1) Different **GreenSKUs** are optimal under different conditions
(2) Using our **GreenSKU** achieves **15**% compute cluster savings at Azure

# Contributions

1. A real **GreenSKU** **server design** and implementation

2. A **framework** to taxonomize the considerations when evaluating a **GreenSKU**'s potential

3. A carbon **evaluation** of our **GreenSKU** with our framework

**See the paper for…**

- a complete outline of **GreenSKU** deployment challenges
- a comparison against other carbon reduction strategies
- an analysis of VM allocation at scale

and more!

Thank you!
Questions?

# Designing Cloud Servers for Lower Carbon

**Jaylen Wang**, Daniel S. Berger, Fiodar Kazhamiaka, Celine Irvene, Chaojie Zhang, Esha Choukse, Kali Frost, Rodrigo Fonseca, Brijesh Warrier, Chetan Bansal, Jonathan Stern, Ricardo Bianchini, Akshitha Sriraman

Carnegie Mellon University

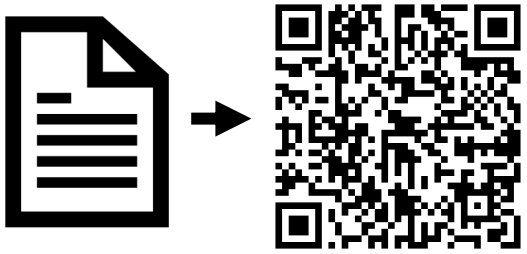
Microsoft Azure

Take a picture to access our **carbon model**


bit.ly/greensku-artifact

ISCA'24

Take a picture to access the **full paper**



bit.ly/greensku-paper