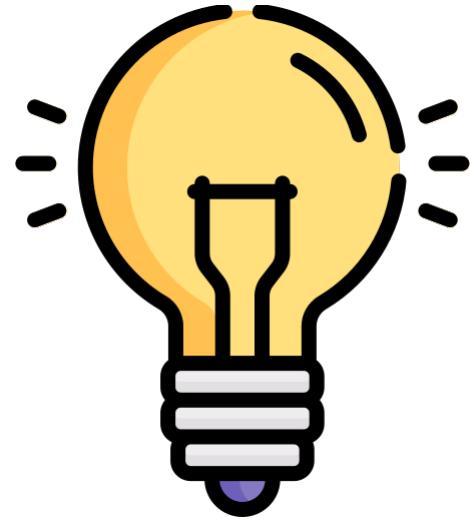


Cost Optimization: How to Actually Do It



What will this
session be about?

Who am I?

@AnnieTalvasto
CMO at VSHN

- CNCF Ambassador
- Azure MVP

- Kubernetes & CNCF meetup co-organizer
- Startup-coach
- Co-host of Cloudgossip podcast - cloudgossip.net



Why does cost
optimization
matter?



Why is cost optimization complex?

a1.medium	\$0.0255	1	2 GiB	EBS Only	Up to 10 Gigabit
a1.large	\$0.051	2	4 GiB	EBS Only	Up to 10 Gigabit
a1.xlarge	\$0.102	4	8 GiB	EBS Only	Up to 10 Gigabit
a1.2xlarge	\$0.204	8	16 GiB	EBS Only	Up to 10 Gigabit
a1.4xlarge	\$0.408	16	32 GiB	EBS Only	Up to 10 Gigabit
a1.metal	\$0.408	16	32 GiB	EBS Only	Up to 10 Gigabit
t4g.nano	\$0.0042	2	0.5 GiB	EBS Only	Up to 5 Gigabit
t4g.micro	\$0.0084	2	1 GiB	EBS Only	Up to 5 Gigabit
t4g.small	\$0.0168	2	2 GiB	EBS Only	Up to 5 Gigabit
t4g.medium	\$0.0336	2	4 GiB	EBS Only	Up to 5 Gigabit
t4g.large	\$0.0672	2	8 GiB	EBS Only	Up to 5 Gigabit
t4g.xlarge	\$0.1344	4	16 GiB	EBS Only	Up to 5 Gigabit
t4g.2xlarge	\$0.2688	8	32 GiB	EBS Only	Up to 5 Gigabit
t3.nano	\$0.0052	2	0.5 GiB	EBS Only	Up to 5 Gigabit
t3.micro	\$0.0104	2	1 GiB	EBS Only	Up to 5 Gigabit
t3.small	\$0.0208	2	2 GiB	EBS Only	Up to 5 Gigabit
t3.medium	\$0.0416	2	4 GiB	EBS Only	Up to 5 Gigabit
t3.large	\$0.0832	2	8 GiB	EBS Only	Up to 5 Gigabit
t3.xlarge	\$0.1664	4	16 GiB	EBS Only	Up to 5 Gigabit
t3.2xlarge	\$0.3328	8	32 GiB	EBS Only	Up to 5 Gigabit

How does cloud pricing work?

- On-demand instances
- Reserved instances (+ Savings plans in AWS)
- Spot VMs/instances
- Dedicated host



How to forecast cloud costs?

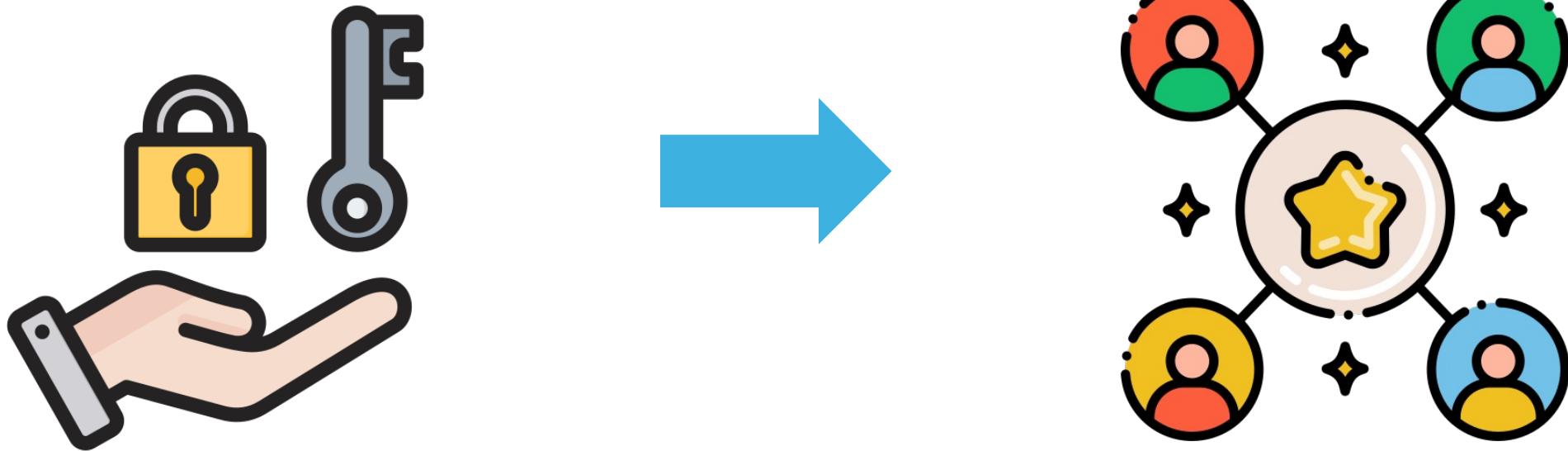
- Analyze your usage reports
- Model your cloud expenses
- Detect peak resource usage scenarios



So many options

- Policies
- Unused or unattached resources
- Idle resources
- Heat maps
- Right sizing
- Reserved instances
- Spot instances
- Multi cloud
- Understand your bill
- Native cost management tools
- Choosing the right compute services
- Autoscaling
- Set budgets, limitations & alerts
- Allocation
- Structure of your resources
- Tagging & labelling
- Review & repeat
- Goal setting
- Dedicated host
- Forecasting

FinOps: from the Old to the New



[Take the State of FinOps 2024 Survey →](#)

Advancing the People of FinOps

The FinOps Foundation is a place for practitioners to advance your career through connections, best practices, education, and specifications.

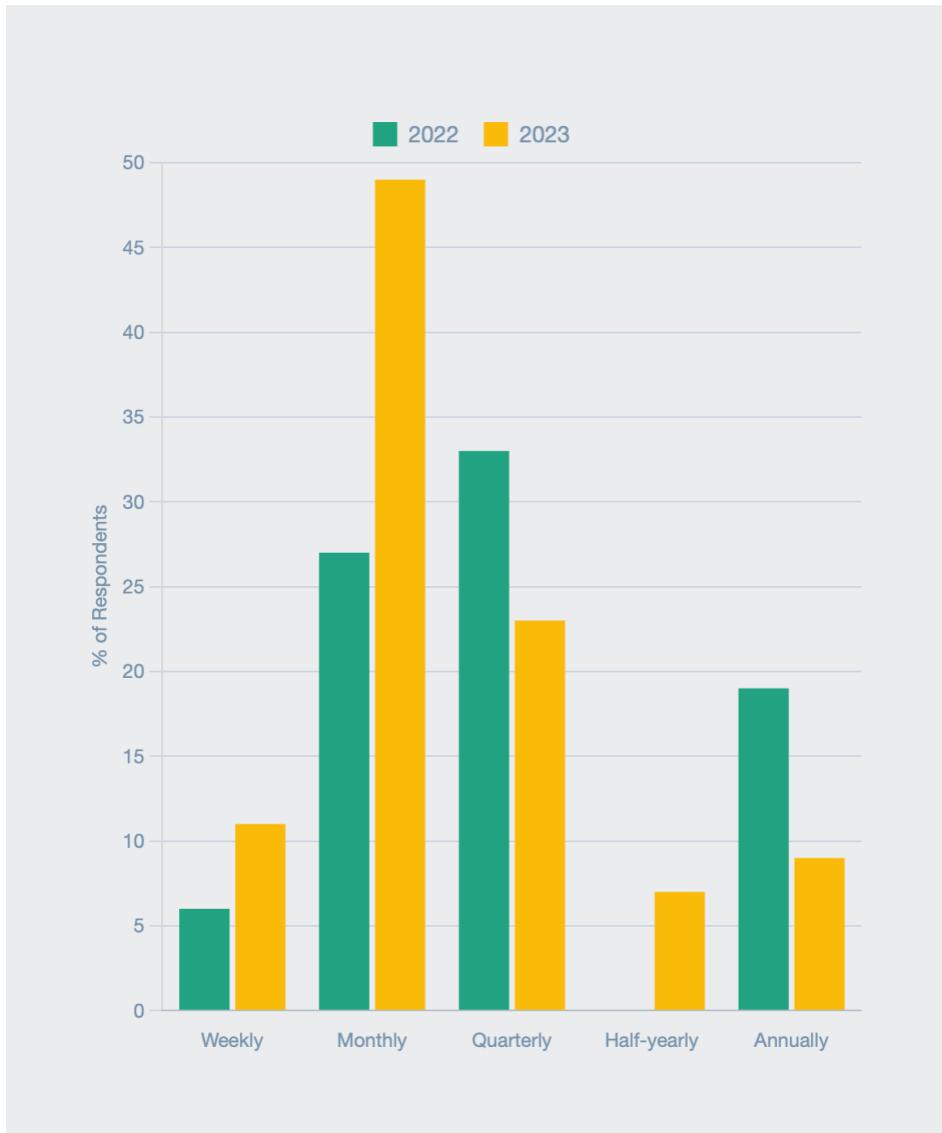
[What is FinOps?](#)

[FinOps Framework Overview](#)

[Training and certification](#)



Forecasting



2022 2023

Forecasting is happening more frequently

Compared to 2022 frequency of forecasting has increased, with 52% decrease in respondents doing annual forecasting and the biggest increase in those forecasting monthly. This indicates this capability is maturing with over 50% showing they are at Walk or Run levels.

Related Framework Capabilities

View maturity information, benchmarks and key performance indicators.

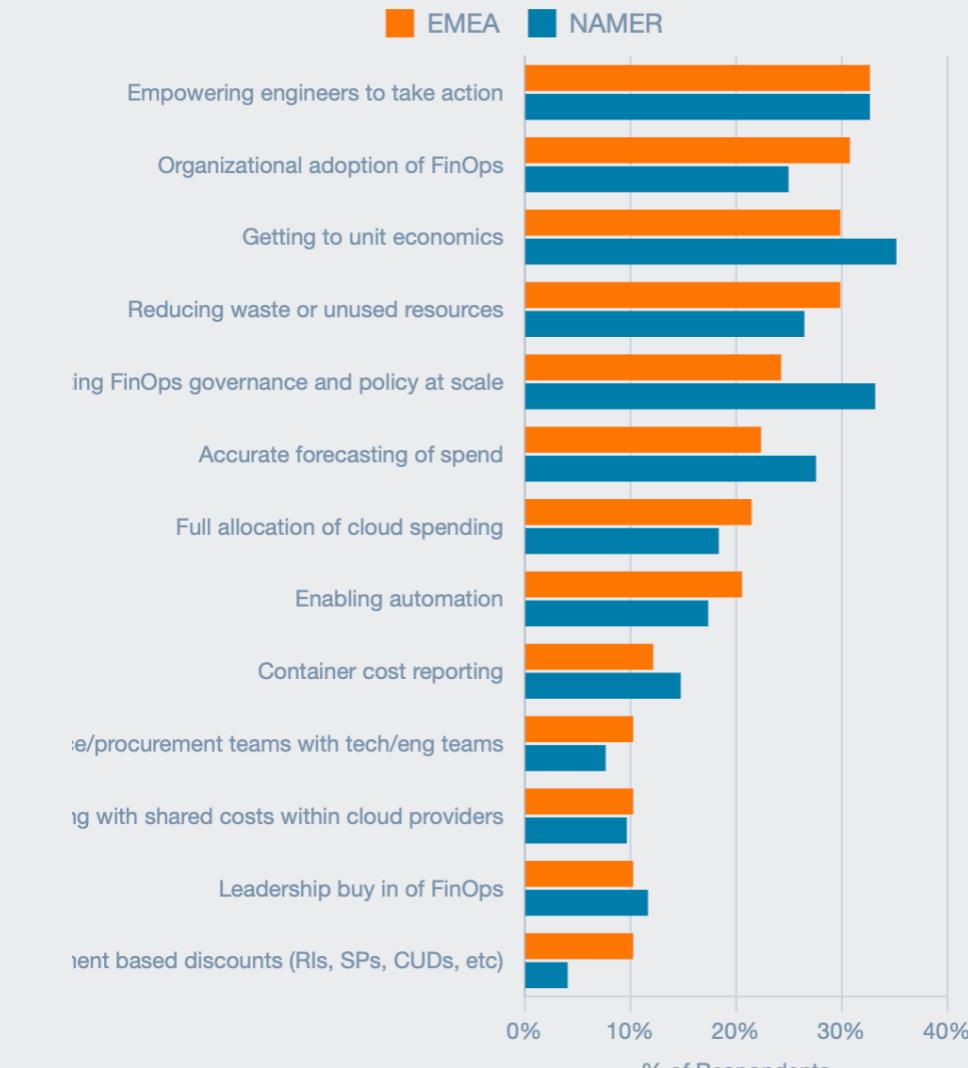
- [Forecasting](#)

Source: FinOps Foundation

[Close](#)

[Copy Link](#)

Top EMEA pain points



What were the top EMEA pain points?

The top swing towards Europe, Middle East and Africa (EMEA) in pain points is around culture. The opposite swings (less pain) are accurate forecasting, unit economics and implementing FinOps governance and policy at scale.

Source: FinOps Foundation

[Close](#)

[Copy Link](#)

Top Challenges

2023



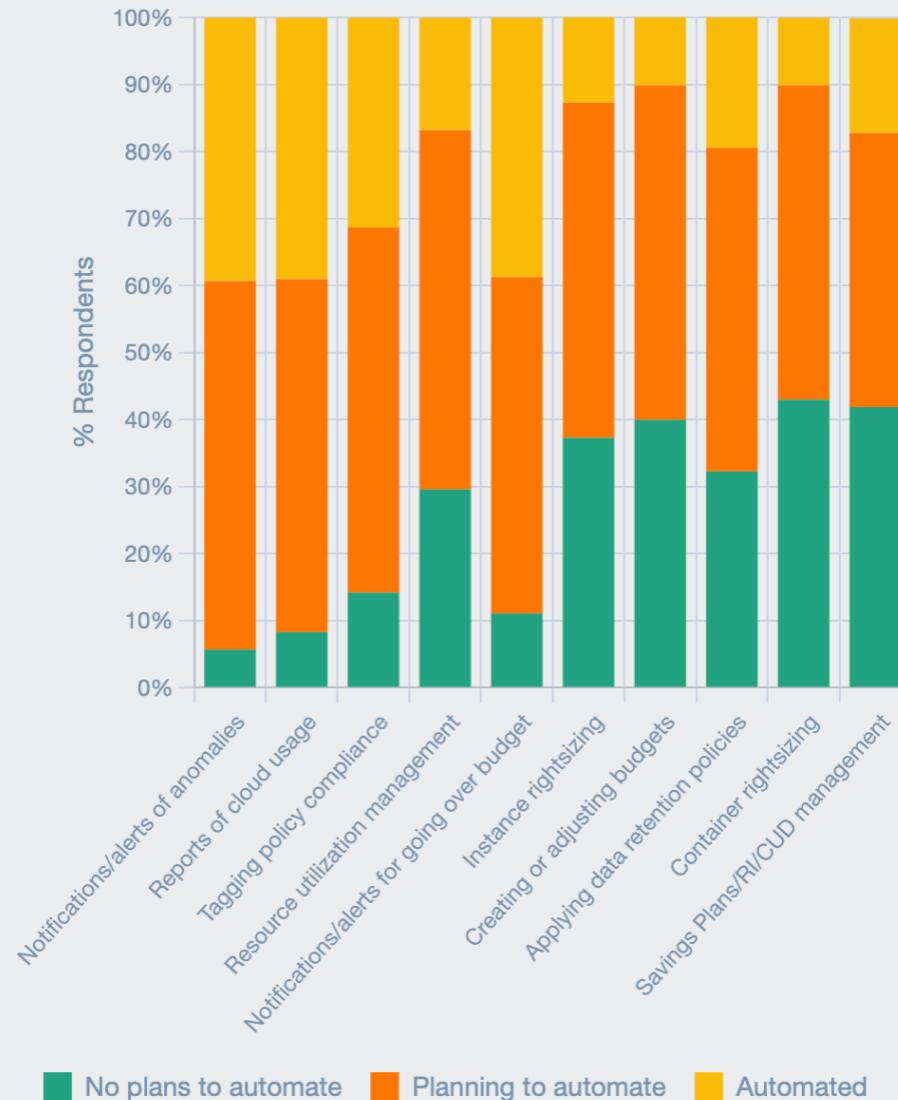
Top challenges for practitioners shift and evolve in 2023

FinOps challenges continue to shift and evolve, changing how practitioners view and prioritize how they tackle them. Compared to 2022, our 2023 respondents still find that Enabling Engineers to Take Action is a top challenge with a minor 0.5% decrease in percentage of respondents. Unit Economics is a new challenge listed this year, and comes in at second position. Organizational Adoption of FinOps remains high from the 2022 and Reducing Waste or Unused Resources percentage increasing 30%. Notably, Forecasting has dropped this year (17%), indicating our respondents increasing their proficiency. Container Costs increases as a notable challenge (15%). *Please note: Arrows represent % directional change vs 2022, multi-select answers.*

Source: FinOps Foundation

@AnnieTalvasto

Automation



2023

The majority are still planning to automate

2023 responses indicate that there's still a long way to go with automation across a lot of areas as, on average, 50% of respondents are still planning to automate. The high amount of respondents who say they have no plans to automate container rightsizing either indicates they're not using containers at all, or they don't see the value in doing so.

Related Framework Capabilities

View maturity information, benchmarks and key performance indicators.

- [Workload Management & Automation](#)

Source: FinOps Foundation

[Close](#)

[Copy Link](#)

FinOps means
working together



Core Ideas

- Cost Transparency

- Monitoring
- Labeling



- Cost control

- Rightsizing
- Managing resources/waste



Cost Transparency: Monitoring

- Performance vs Cost



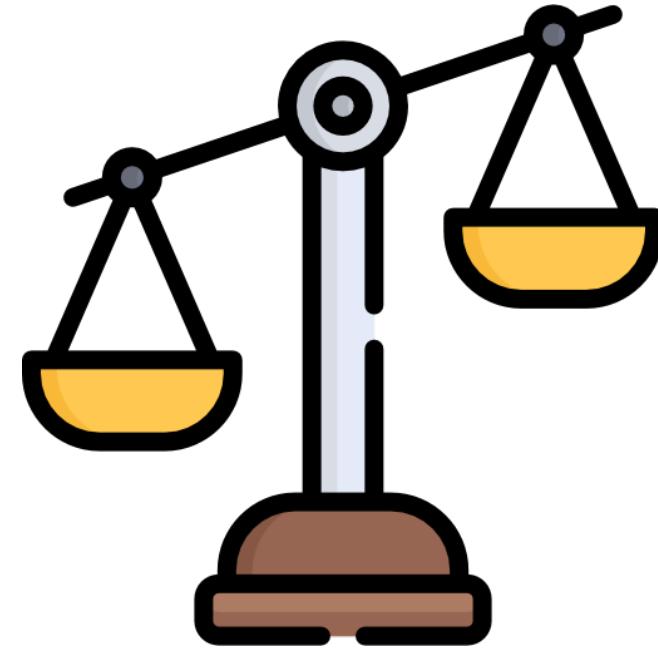
Cost Transparency: Labeling

- Must be understandable for business

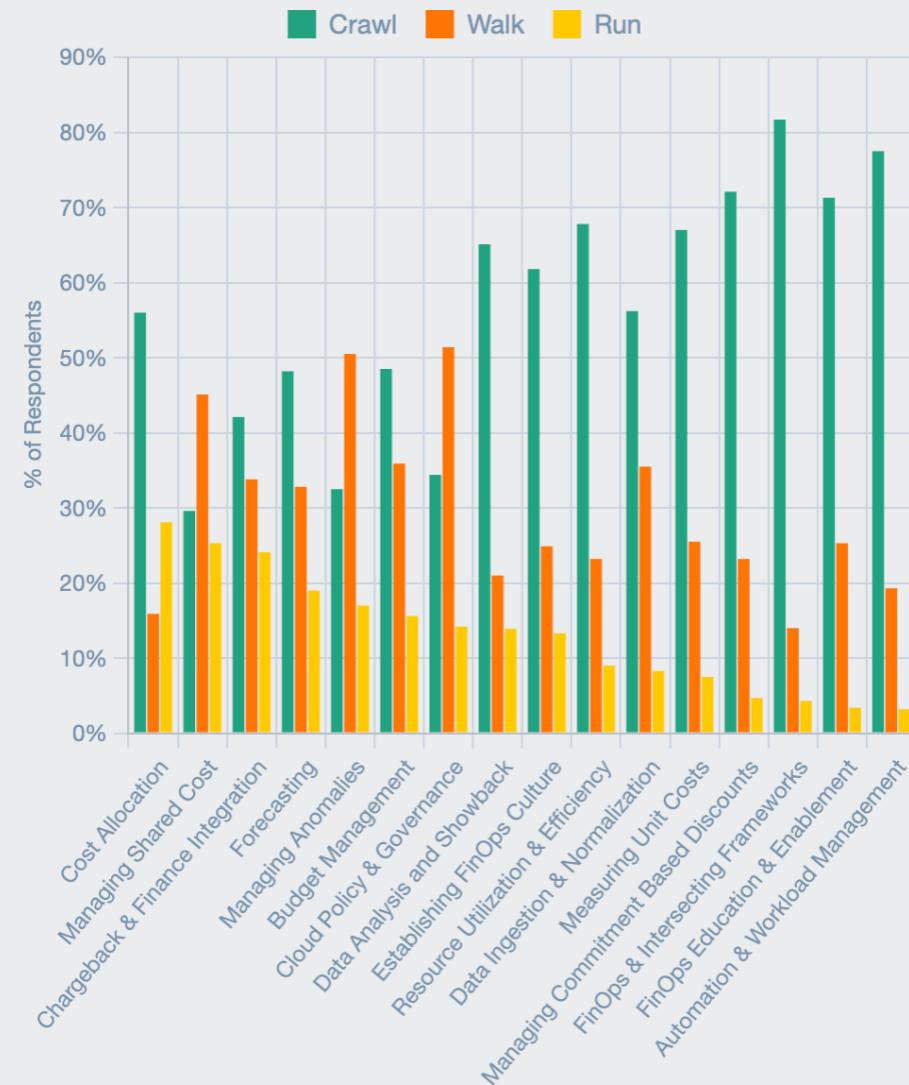


Cost Control: Rightsizing

- VPA (Vertical Pod Autoscaler)
- HPA (Horizontal Pod Autoscaler)
- Cluster Autoscaler
- KEDA (Kubernetes-based Event-Driven Autoscaler)



Rightsizing



Respondents are strong in Cost Allocation, room for improvement with Rightsizing and Automation

We measured respondents by their benchmarks and sorted them by maturity level. In 2023, practitioners are best at Cost Allocation and activities related to Chargeback or Showback. Our more mature cohort shows strength in Forecasting and Managing Anomalies. Automation and Resource Utilization & Efficiency seem to be heavily worked on by our less experienced cohort.

Source: FinOps Foundation

[Close](#)

[Copy Link](#)

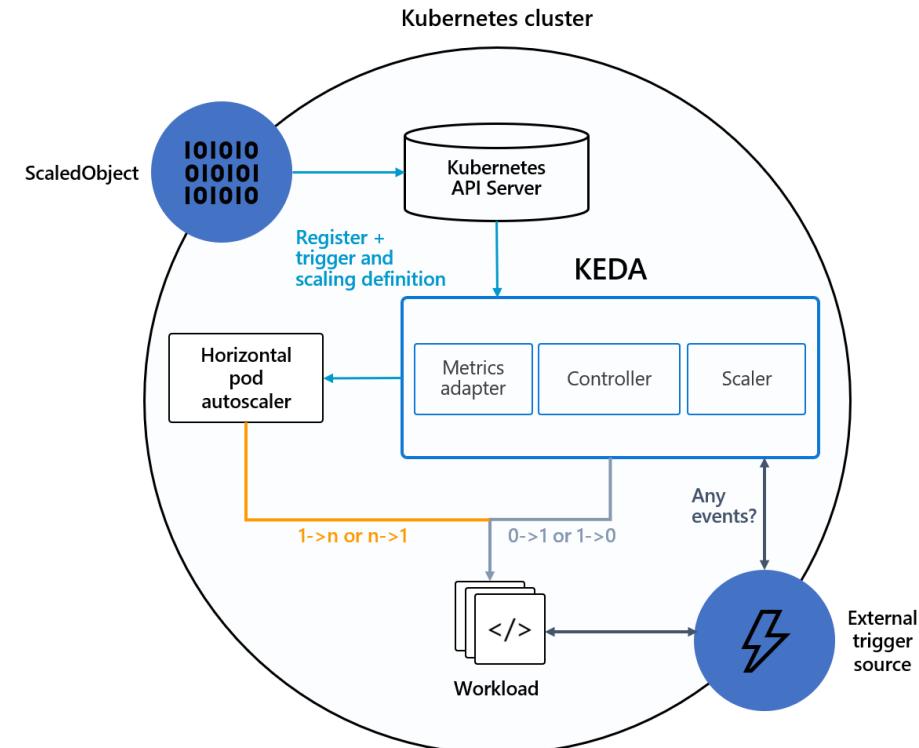
Keda

Kubernetes Event-driven
Autoscaling. Application
autoscaling made simple.



What is Keda?

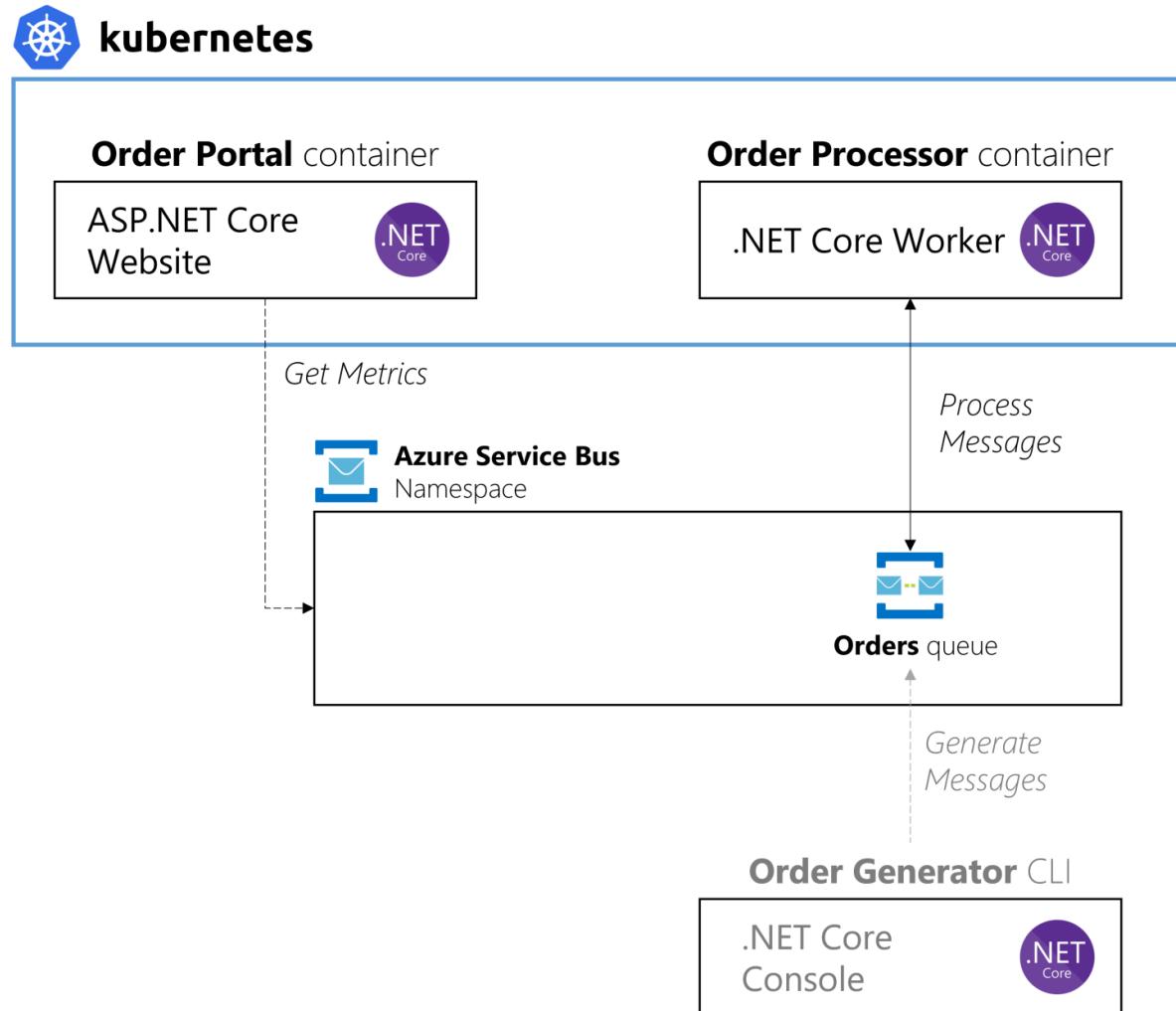
- Default Kubernetes Scaling is not well suited for event driven applications, kubernetes is more for resource based scaling (CPU and memory).
- Keda: Event driven scale controlling that can run inside any kubernetes cluster.



Source: <https://keda.sh/docs/2.10/concepts/>

Demo

KEDA:
Scaling .NET
Core worker
with Azure
Service Bus



Cost Control: Eliminating waste

- Policies



8 best practices for cost optimization



- Define your requirements
- CPU
- Memory
- Storage
- Network

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- A lot of instances types to choose from

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Each application out there comes with its unique storage needs. Make sure that the VM you pick has the storage throughput your workloads need.

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- How much time does your workload need to finish the job?
- Is it mission- and time-critical?
- Can it handle interruptions?
- Is it tightly coupled between instance nodes?
- What tools are you going to use to move your workload when cloud provider pulls the plug?

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- Cherry-pick spot instances
- Once you pick an instance, check its frequency of interruption – the rate at which this instance reclaimed capacity during the trailing month.
- Cloud service Advisors help (Azure Advisor, AWS Spot Instance Advisor)

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- Cherry-pick spot instances
- Bid your price on spot
- Here's a rule of thumb: Set the maximum price to one that equals the on-demand price.

8 best practices for cost optimization



- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- Cherry-pick spot instances
- Bid your price on spot
- Use mixed instances
- Use A mixed-instance strategy gets you great availability and performance at a reasonable cost.

8 best practices for cost optimization

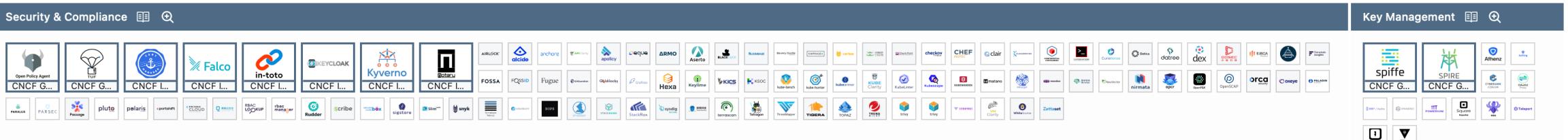
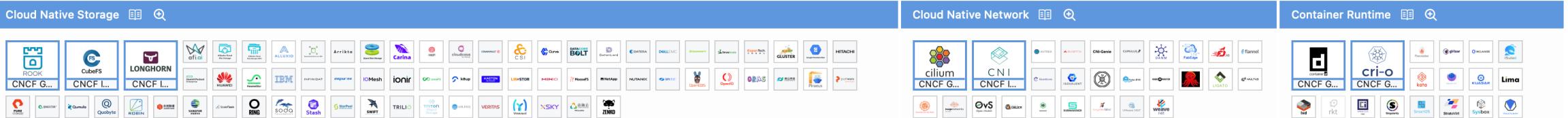
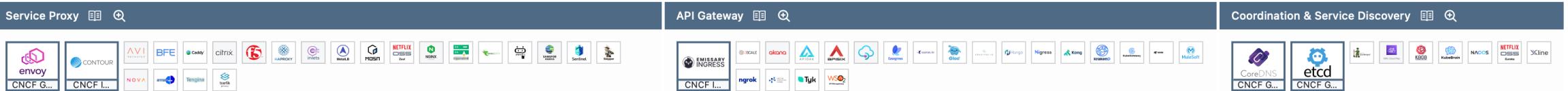
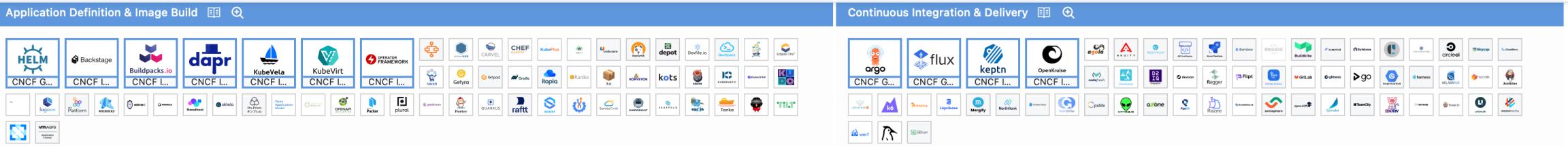


- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- Cherry-pick spot instances
- Bid your price on spot
- Use mixed instances
- Make multiple regions work for you
- Configure multiple node groups
- Scope each of them to a single cloud region, automate your deployments to take advantage of price fluctuations between regions.

8 best practices for cost optimization

- Define your requirements
- Choose the right instance types
- Verify storage transfer limitations
- Check if your workload is spot-ready
- Cherry-pick spot instances
- Bid your price on spot
- Use mixed instances
- Make multiple regions work for you





Vendor Solutions

- Examples:
- CAST AI
- Stormforge
- Spot.io
- Kubecost
- And manymore



Consulting Option

- Cost Optimization Masterclass:



The purpose of all of
this:

Better management
and visibility



Recap

- Why cost optimization
- The basics of cost & forecasting
- FinOps foundation
- Core ideas of cost optimization
- KEDA
- 8 best practices
- CNCF projects, vendors, and consulting



Learn more

- FinOps foundation principles
- <https://www.finops.org/framework/principles/>

- 8 best practices to reduce Kubernetes cost
- <https://konghq.com/blog/cloud-cost-optimization/>

- FinOps Foundation principles
<https://www.finops.org/framework/principles/>

- CNCF & FinOps Youtube
<https://www.youtube.com/@cncf/>
<https://www.youtube.com/@finopsfoundation/>

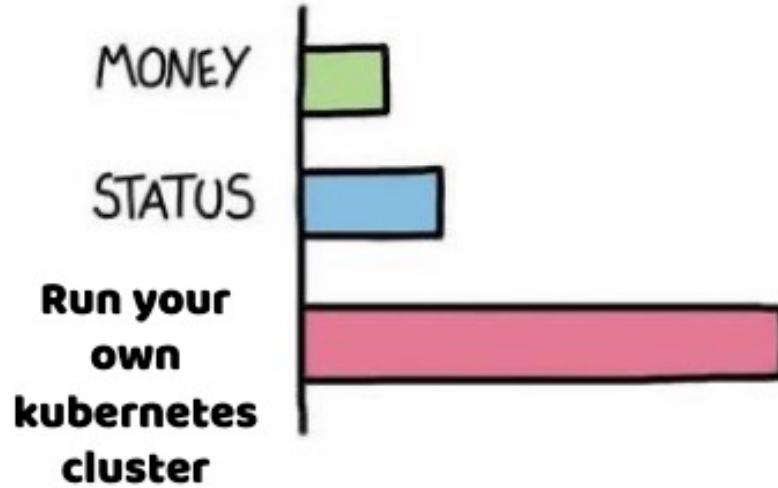
- Links and slides: github.com/annietalvasto

Cost Optimization
Masterclass:



WHAT GIVES PEOPLE FEELINGS OF POWER

Thank you!



@iamnotanerd...

@AnnieTalvasto