



# Az COMMUNITY

—— Conference 2022 ——

Asia's Largest Azure Community Conference

#AzConfDev

# Design for costs – Four pillars of economic cloud design

Michiel Hamers & Twan Koot – Ordina



#AzConfDev

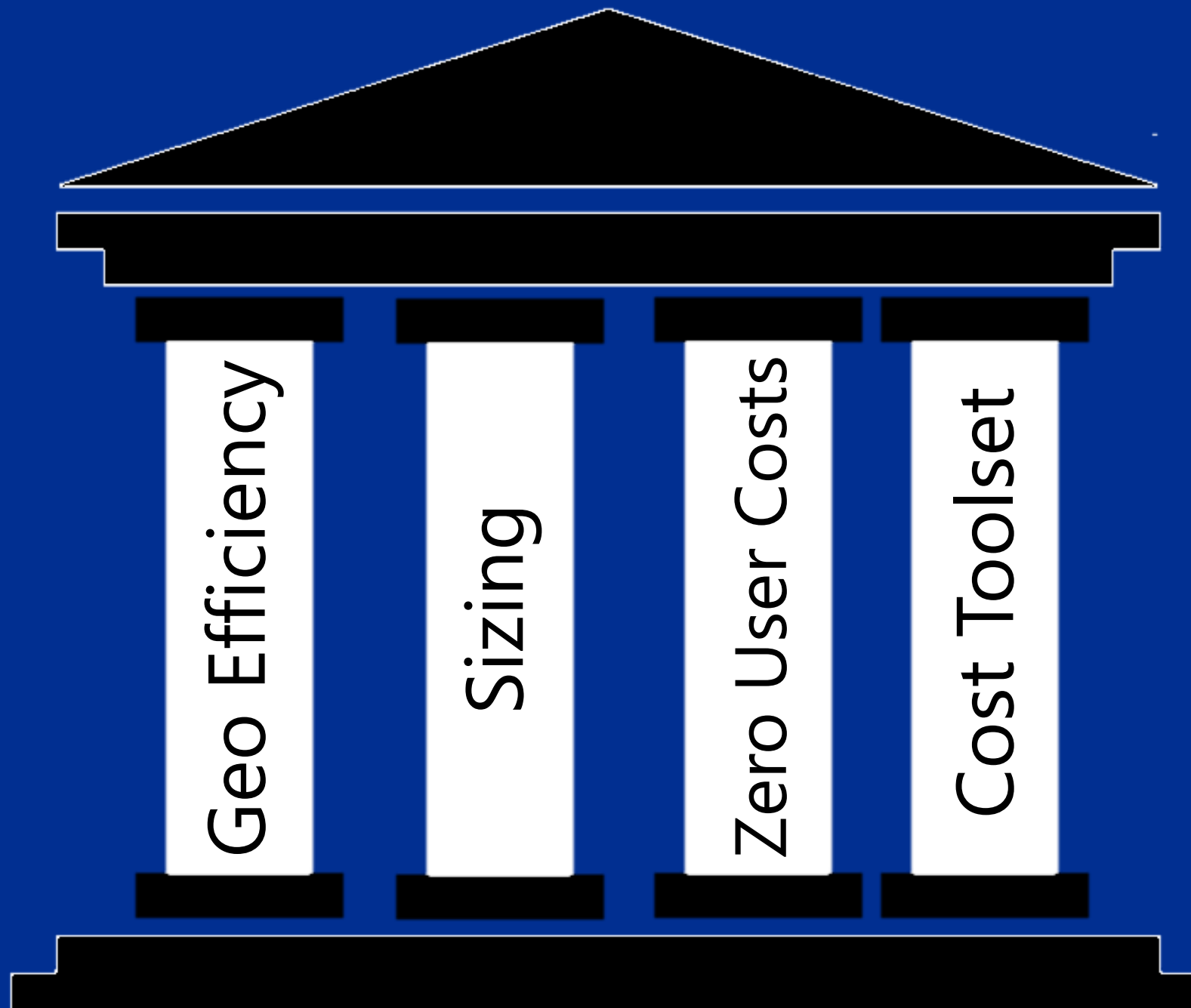
# Agenda

Introduction

Why?

4 pillars

Tips, Tricks & QA



#AzConfDev



# Michiel Hamers

---

Solution Lead Microsoft (Azure) Cloud &  
Codesmith - Ordina



#AzConfDev



# Twan Koot

---

Lead performance engineer &  
Codesmith - Ordina



#AzConfDev

# Disclaimer



All prices mentioned in this presentation are subject to change. Pricing may vary for each azure user based on their agreement, usage and local currency rates.

Any mentioned costs savings are based on the public azure calculator prices or real life examples from Ordina's Azure environment.



# Why?



# Geo Efficiency

The first pillar



#AzConfDev



# Geo Efficiency - Geo what ??



"Geo efficiency is the art of getting the same amount compute power at a different region for a lower price"



#AzConfDev

# Geo Efficiency - It's all about the money

There is a big difference in datacentres pricing between each region.  
Let's select a VM : D16d v4 - 16 vCPU 64GB RAM OS:Ubuntu

Germany North: ₹84,225.75

North: Europe ₹60,297.97

Switzerland west: ₹93,080.20

Brazil SouthEast: ₹112,041.76

South Africa North: ₹71,783.29

South India: ₹76,209.93



# Geo Efficiency - Impact at scale

One of Ordina's customers is running an azure environment with a bill of \$4.000.000,00 each month.



#AzConfDev

# Geo Efficiency - Picking an Azure region



## Choose the right Azure region for you



### Compliance and data residency

Get help choosing the right geography for your residency and compliance needs.

[Learn more about data residency](#)

[See compliance offerings](#)



### Service availability

Ensure that the Azure services you need are available in the datacenter region that you're considering.

[Explore products by region](#)



### Pricing

Factor cost into your decision-making process.

[Explore pricing](#)



#AzConfDev

# Geo Efficiency - Let's compare



Let's select a VM : D16d v4 - 16 vCPU 64GB RAM OS:Ubuntu

South India: ₹76,226.32 ~23% Difference Central India: ₹58,400.14

Now let's apply the 1 year reserve 41% discount:

Central India: ₹34,327.97

This means a potential save of 50%!!



#AzConfDev

# Geo Efficiency - Latency impact

Let's compare Azure South vs Central India



We deployed a Website using appservice using a P1V2 tier as App service plan:

- Running HTTP 2.0
- Disabled ARR
- .NET core running windows



#AzConfDev

# Geo Efficiency - Latency impact

Let's compare Azure South vs Central India



ping d4psouthindia.southindia.cloudapp.azure.com -n 100  
ping d4pcentralindia.centralindia.cloudapp.azure.com -n 100

count	Avg(ms)	Min(ms)	Max(ms)
100	2	2	4
100	29	28	34



# Geo Efficiency - Latency impact

Let's compare Azure South vs Central India

Running a jmeter 5 thread performance test fetching a demo website including all embedded recourses(7 HTTP GET's).

Region	Test runs	Avg(ms)	Std. Dev.(ms)	Min(ms)	TPS
South	1000	430	78	285	11,2/Sec
Central	1000	550	84	418	8,1/Sec





# Geo Efficiency - What's the impact?

How much impact do you have when switching from south to central India?



- Latency to the datacentre
- Availability in services
- Key-vault entries\*
- Public IP's
- Webhook triggers



- ACU / Plan
- BICEP and ARM templates\*
- Settings
- Tags\*
- Policies



# Geo Efficiency - 4 steps in region selection



Following Azure's guidelines deploying a new azure resources should follow the following these steps:

- 1.) Choose the region based on data residency/compliance
- 2.) Discover which region support your desired service
- 3.) Determine the network performance needs of your service
- 4.) Select the most economical region fitting your needs



# Geo Efficiency - Dev/Test



Deploying Dev/Test can also follow the previous approach:

- For Dev/Test environments using the cheapest option is recommend
- Running your functional test against a slower environment is perfectly fine
- For non functional testing(Performance) it can cause different performance metrics



#AzConfDev



# Sizing

The second pillar



#AzConfDev



# Sizing

Tiers\*

Free  
Shared  
Basic  
Standard



#AzConfDev



\*Tiers like basic are obsolete and should not be used anymore.

# Sizing VMs



Type	Sizes	Description
General purpose	B, Dsv3, Dv3, Dasv4, Dav4, DSv2, Dv2, Av2, DC, DCv2, Dpds5, Dplds5, Dpsv5, Dplsv5, Dv4, Dsv4, Ddv4, Ddsv4, Dv5, Dsv5, Ddv5, Ddsv5, Dasv5, Dadsv5	Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers.
Compute optimized	F, Fs, Fsv2, FX	High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers.
Memory optimized	Esv3, Ev3, Easv4, Eav4, Epds5, Epsv5, Ev4, Esv4, Edv4, Edsv4, Ev5, Esv5, Edv5, Edsv5, Easv5, Eadsv5, Mv2, M, DSv2, Dv2	High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.
Storage optimized	Lsv2, Lsv3, Lasv3	High disk throughput and IO ideal for Big Data, SQL, NoSQL databases, data warehousing and large transactional databases.
GPU	NC, NCv2, NCv3, NCasT4_v3, ND, NDv2, NV, NVv3, NVv4, NDasrA100_v4, NDm_A100_v4	Specialized virtual machines targeted for heavy graphic rendering and video editing, as well as model training and inferencing (ND) with deep learning. Available with single or multiple GPUs.
High performance compute	HB, HBv2, HBv3, HC, H	Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA).



#AzConfDev

# Sizing VMs



Azure  
Compute  
Unit

(ACU)\*

SKU Family	ACU \ vCPU	vCPU: Core
A1_v2 - A8_v2	100	1:1
A2m_v2 - A8m_v2	100	1:1
B	Varies	1:1
D1 - D14	160 - 250	1:1
D1_v2 - D15_v2	210 - 250*	1:1
DS1 - DS14	160 - 250	1:1
DS1_v2 - DS15_v2	210 - 250*	1:1
D_v3	160 - 190*	2:1***
Ds_v2	160 - 190*	2:1***

\*The ACU is only a guideline. The results for your workload may vary.



# Sizing - Naming



**[Family]** + *[Sub-family]\** + **[# of vCPUs]** + *[Constrained vCPUs]\** + **[Additive Features]** + *[Accelerator Type]\** + **[Version]**

Example: M8-2ms\_v2 (Constrained vCPU)

Value	Explanation
Family	M
# of vCPUs	8
# of constrained (actual) vCPUs	2
Additive Features	m = memory intensive s = Premium Storage capable
Version	v2

Value	Explanation
Family	Indicates the VM Family Series
*Sub-family	Used for specialized VM differentiations only
# of vCPUs	Denotes the number of vCPUs of the VM
*Constrained vCPUs	Used for certain VM sizes only. Denotes the number of vCPUs for the <a href="#">constrained vCPU capable size</a>
Additive Features	One or more lower case letters denote additive features, such as: a = AMD-based processor b = Block Storage performance c = confidential d = diskful (i.e., a local temp disk is present); this is for newer Azure VMs, see <a href="#">Ddv4</a> and <a href="#">Ddsv4-series</a> i = isolated size l = low memory; a lower amount of memory than the memory intensive size m = memory intensive; the most amount of memory in a particular size t = tiny memory; the smallest amount of memory in a particular size s = Premium Storage capable, including possible use of <a href="#">Ultra SSD</a> (Note: some newer sizes without the attribute of s can still support Premium Storage e.g. M128, M64, etc.) NP = node packing p = ARM Cpu
*Accelerator Type	Denotes the type of hardware accelerator in the specialized/GPU SKUs. Only the new specialized/GPU SKUs launched from Q3 2020 will have the hardware accelerator in the name.
Version	Denotes the version of the VM Family Series





# Sizing - Generation switching

A offer you can't refuse.

VM Name	vCPUs	Memory (GiB)	Linux Cost	Windows Cost
Standard_D2	2	7	135670.50	201711.26
Standard_D2_v2	2	7	121313.74	187354.50
Standard_D2_v3	2	8	96907.50	170126.21
Standard_D2_v4	2	8	96907.50	162948.26
Standard_D2_v5	2	8	96907.50	162948.26

Example: INR, "per year", datacenter Chennai, standard prices, Pay as-you-go



# Sizing - Azure advisor



Azure advisor is for VMs

Recommendation Status : **Active**

Add Filter

Reserved Instances Configuration : **1 years, 60 days**



Total recommendations

4

Recommendations by impact

4

High  
impact

0

Medium  
impact

0

Low  
impact

Impacted resources

3



Potential yearly savings based on retail pricing

50,004 USD



For more cost management and optimization capabilities, visit [Cost Management](#)

Impact

Description

Potential yearly savings based on retail pricing

Impacted resources

Last updated

High

Buy virtual machine reserved instances to save money over pay-as-you-go costs

48,155.01 USD

1 Virtual machine

17/10/2022, 21:34

High

Consider Red Hat reserved instance to save over your on-demand costs

421.00 USD

1 Subscription

17/10/2022, 11:57

High

Consider Database for PostgreSQL reserved instance to save over your pay-as-you-go costs

648.00 USD

1 Subscription

17/10/2022, 13:42

High

Right-size or shutdown underutilized virtual machines

[Quick fix](#)



780.00 USD

2 Virtual machines

17/10/2022, 17:05

# Sizing - Monitoring advisor



Select	Virtual machine	Recommended actions	Potential yearly savings based on di...	Subscription	Recommendation rule	Additional details**	Last updated	Action
<input type="checkbox"/>	 vm-bmc-patrol-rt02	Shut down the virtual machine <a href="#">View Usage Patterns</a>	<b>107.14 USD</b>	Ordina NL BPS - Remote Management	CPU utilization < 5%	CPU (%) - 3 Network (%) - 0 Memory (%) - 70	17/10/2022, 17:05	<a href="#">Postpone</a>   <a href="#">Dismiss</a>
<input type="checkbox"/>	 avail-vm-mgthub-prod-asa-b	Resize Standard_D3_v2 to Standard_F4s_v2 <a href="#">View Usage Patterns</a>	<b>696.38 USD (29%)</b>	Ordina NL BPS - Remote Management	CPU utilization < 5%	CPU (%) - 3 Network (%) - 0 Memory (%) - 16	17/10/2022, 16:49	<a href="#">Postpone</a>   <a href="#">Dismiss</a>



#AzConfDev

# Sizing - Prices

- Listed prices
- Personal prices
- Discounts



#AzConfDev



Not equal to

# Sizing - Scaling



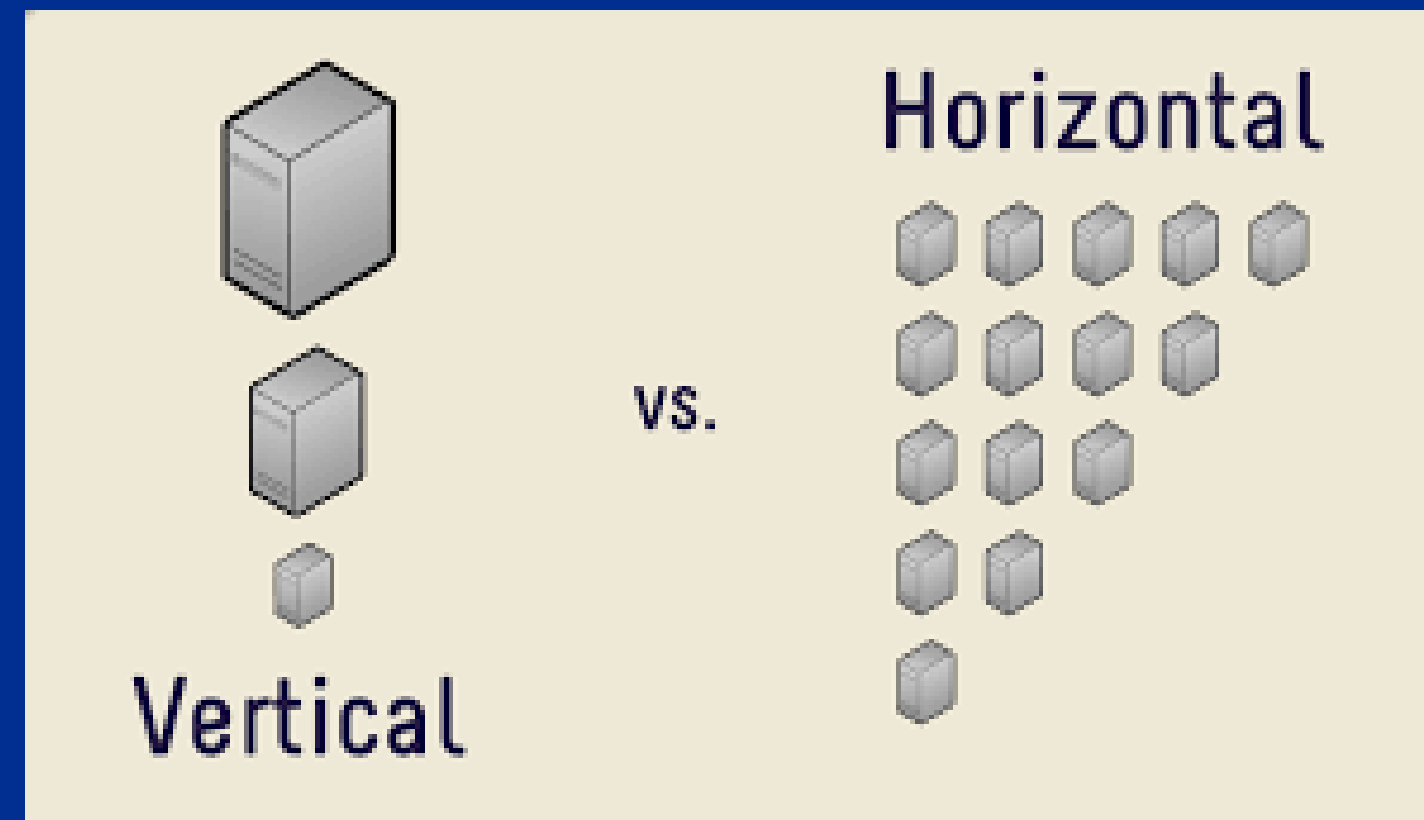
Scaling -> Increasing the size or count of servers/computing power used

Auto -> the automatic process of scaling

Tip: Use Horizontal scaling (scale out/in)!



#AzConfDev



# Sizing - Picking the right size



How do you pick the right size for your application or services:

- Take advantages of azure's capabilities -> resize when needed
- Performance testing is 100% needed in order to determine the right size
- Smaller instances allow for:
  - **Fine grained control in scaling**
  - **Better throughput of requests**



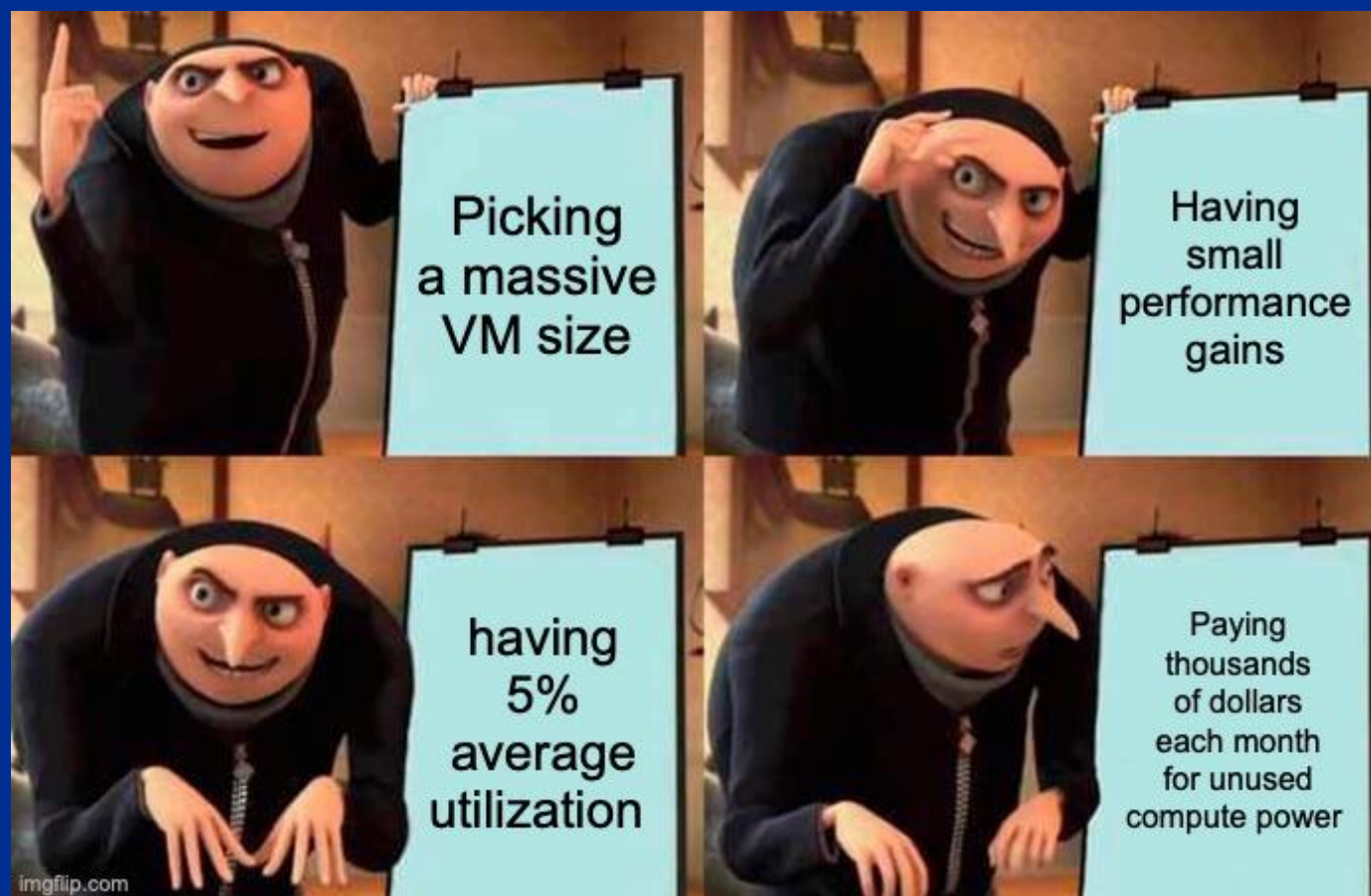


# Sizing - Picking the right size



Controversial statement incoming:

"Having 100% CPU and Memory utilization is fine! I'm using everything I'm paying for"



#AzConfDev

# Sizing - Picking the right size

- Look at response times instead of utilization
- Running daily load between 60-75% utilization should be the goal



#AzConfDev

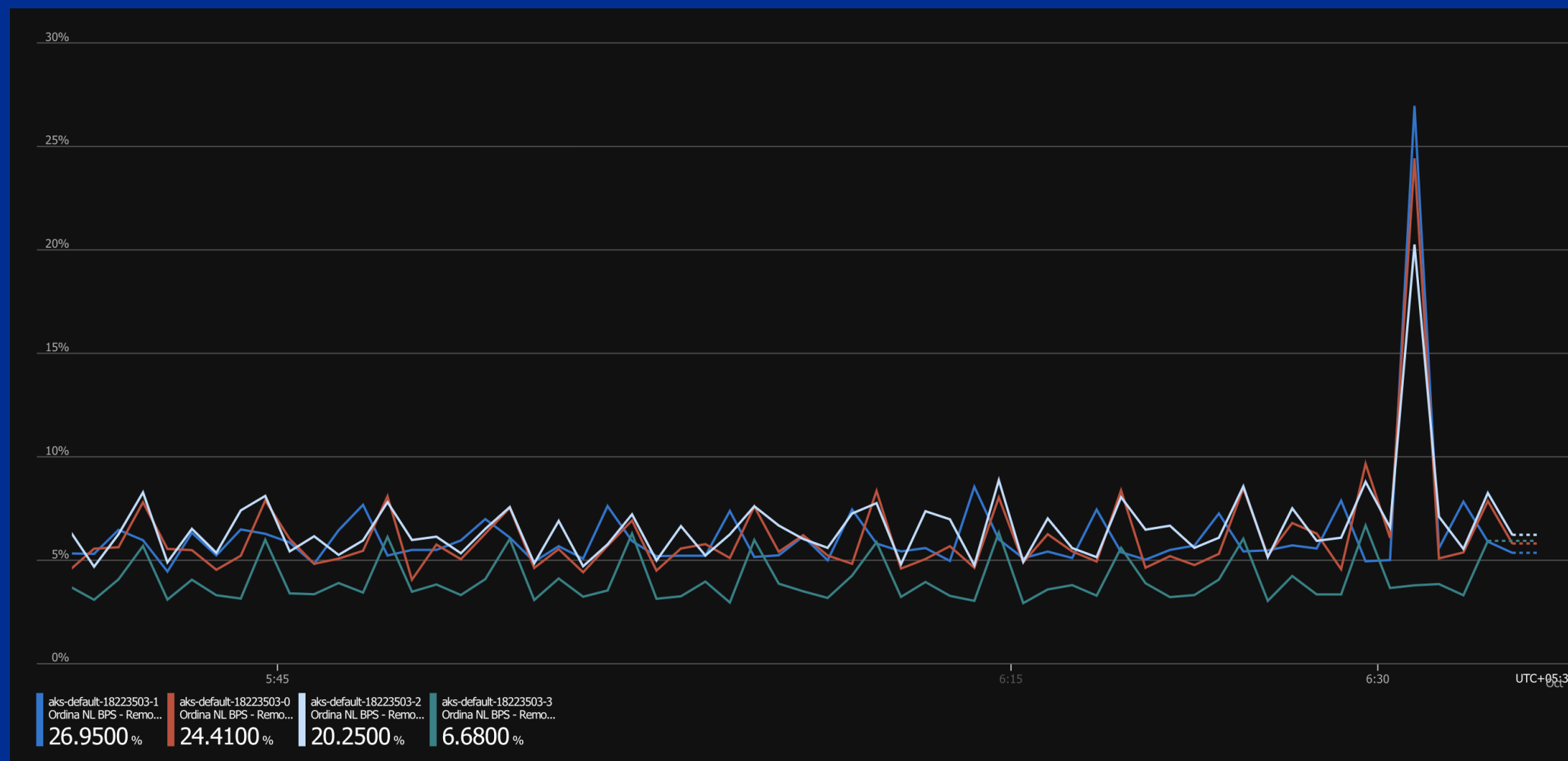




# Sizing - Monitoring



Keep an eye on your monitoring for optimization so you know when to use scale, or switch generation hardware.



#AzConfDev

# Zero User Costs

The third pillar



#AzConfDev

# Zero User Costs - Let's Explain



“the amount of costs associated with running your cloud environment without any user based activity”



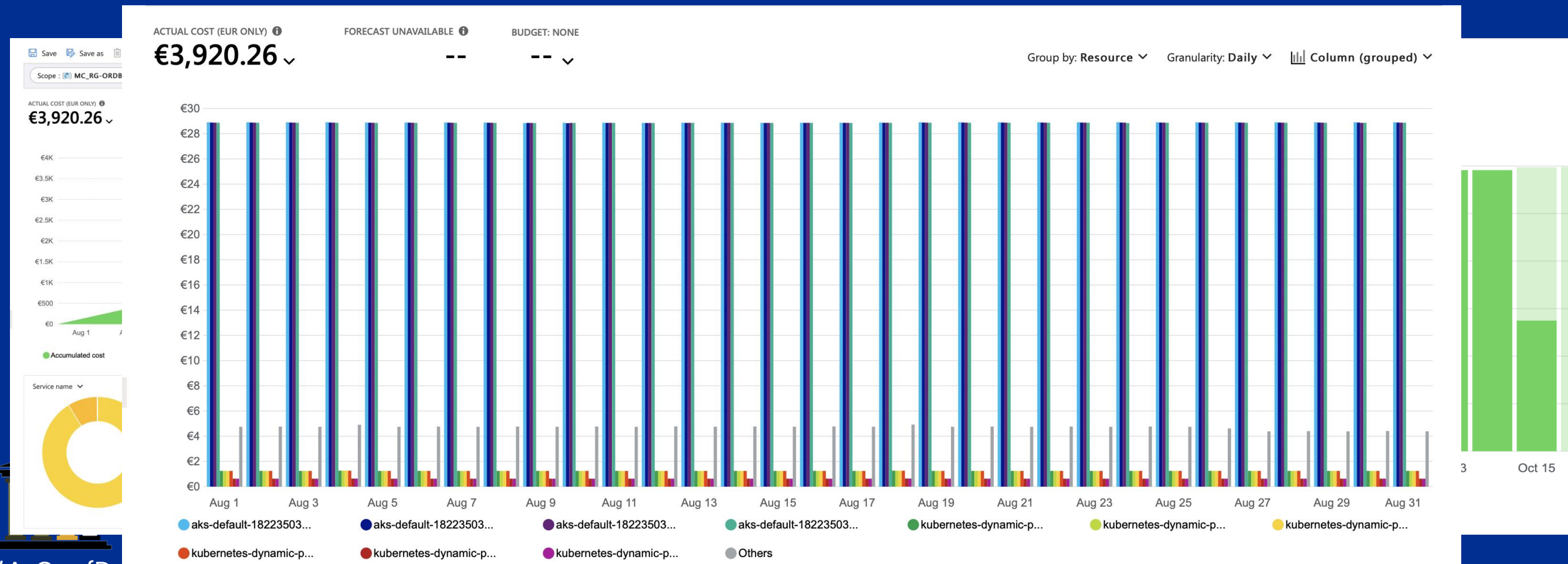
#AzConfDev

# Zero User Costs - Let's Explain



How can you spot zero user costs in your Azure env. ?

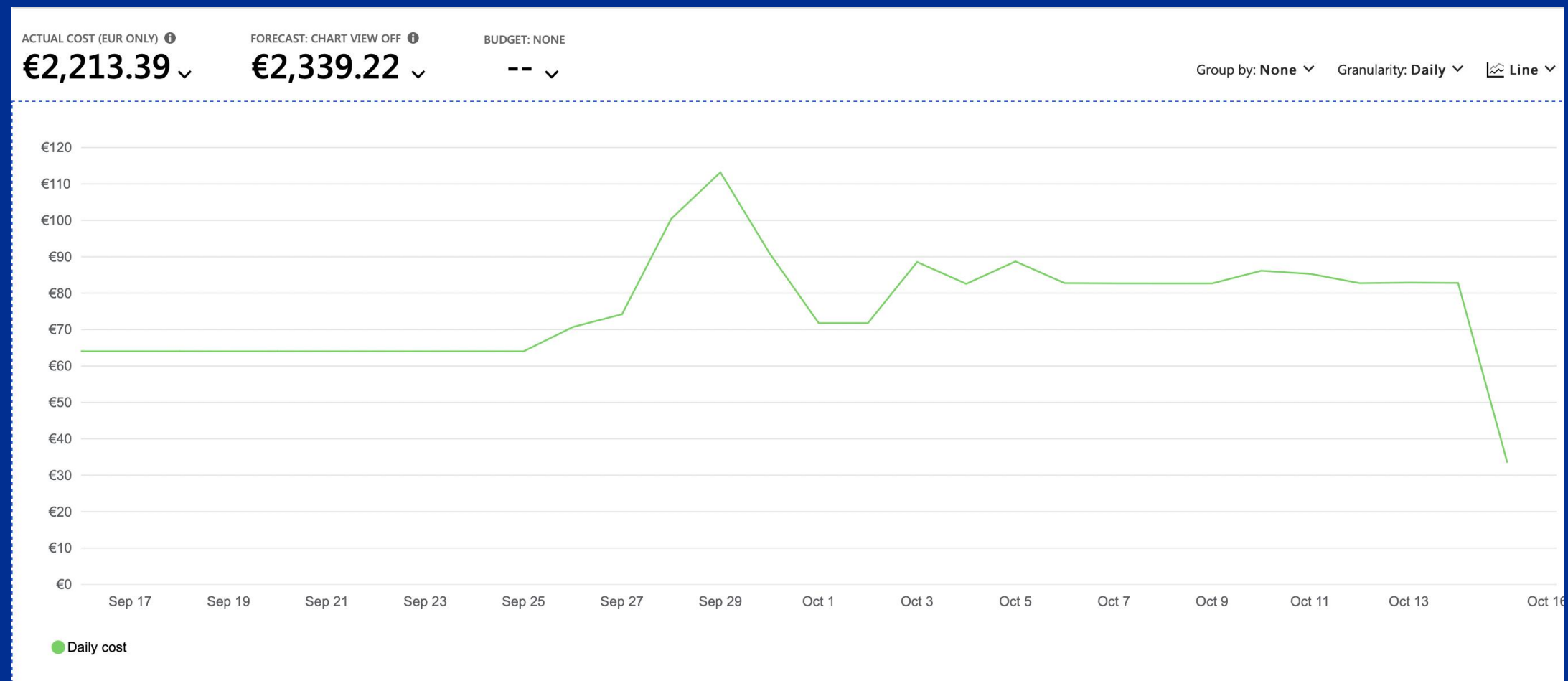
Well linear increasing costs is likely to be contributing to your zero user costs



# Zero User Costs



To achieve a good zero user costs your azure costs graph should look like ->



#AzConfDev

# Zero User Costs - Why?



Why would you want to monitor zero user costs ?

- Create visibility in your fixed cloud costs
- Determine the level of cloud native development adoption
- Having a baseline helps in tracking improvement or decline
- Finance/costs are a part of DevOps teams and they should be responsible for it



#AzConfDev

# Zero User Costs - Apply these patterns



What is a pattern?

*"Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice" (Alexander et al., 1977).*



#AzConfDev

# Zero User Costs - Anti Patterns



What is a anti pattern?

*"An anti-pattern in software engineering, project management, and business processes is a common response to a recurring problem that is usually ineffective and risks being highly counterproductive." (Koenig, Andrew (1995). "Patterns and Antipatterns". Journal of Object-Oriented Programming).*

Golden Hammer



Copy and Paste Programming



Swiss knife Object



#AzConfDev



# Zero User Costs - Golden Hammer

- Don't assume the previous solution is the right fit for you current problem!
- Fixing performance issues is always a unique approach for each env.
- Having costs as one key deciders in a solution helps to reduce zero user costs



# Zero User Costs - Copy and Paste Programming

- Copying Biceps or ARC templates is all fine and can help in speeding up deployments
- Only use the core of the template and fill in these parts based on your env:
  - Parameters, SKU size, auto scaling settings, storage settings, run book automation



#AzConfDev

# Zero User Costs - God Object/ Swiss Army Knife



- Having a monolithic application = higher zero user costs
- Applying Cloud native architecture is the road towards zero user costs
- Dividing your application across different compute units in azure allows for better implementation of:
  - Scaling down under utilized components
  - Picking the right instance size
  - Decoupling and using PAAS solutions



# Zero User Costs - How to reduce these costs?



How do you lower your zero user costs:

- Use azure reserved pricing or savings plan\*
- Use Dev/test plans on development and testing env
- Use scaling for your production environments
- When using licensed infrastructure or software when applicable use BYOL
- Use performance testing to find your optimal daily run utilization



#AzConfDev



# Cost Toolset

The fourth pillar



#AzConfDev



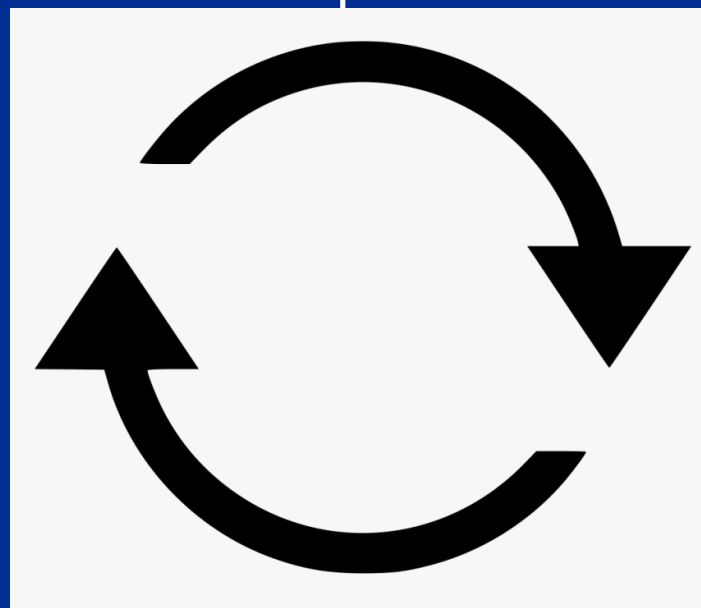
# Cost Toolset



Improve

SMALL  
STEPS  
EVERY  
DAY

Repeat



Reuse



#AzConfDev

# Cost Toolset - (Re)use



When to use Hot/Warm/Cold data path



Path	Requirement
Hot data path	<ul style="list-style-type: none"><li>• When data requirements are known to change frequently</li><li>• When processing or displaying data in real time</li></ul>
Warm data path	<ul style="list-style-type: none"><li>• When you need to store or display a recent subset of data</li><li>• Used for data that is consumed for small analytical and batch processing</li></ul>
Cold data path	<ul style="list-style-type: none"><li>• When data is rarely used. The data might be stored for compliance or legal reasons</li><li>• Used for data that is consumed for long term analytics and batch processing</li></ul>



# Cost Toolset - (Re)use



When to use Hot/Warm/Cold data path



## Azure Storage Reserved Capacity

Azure Storage Reserved Capacity helps you lower your data storage cost by committing to one-year or three-years of Azure Storage. Reserved capacity can be purchased in increments of 100 TB and 1 PB sizes for 1-year and 3-year commitment duration. All prices are per month. For more information, please see [documentation](#).

	1-year reserved			3-year reserved		
	Hot	Cool	Archive	Hot	Cool	Archive
100 TB / month	₹137,612	₹75,690	₹14,935	₹110,762	₹68,303	₹13,763
1 PB / month	₹1,340,427	₹737,234	₹146,073	₹1,065,469	₹661,622	₹134,043





# Cost Toolset - (Re)use

## Azure reservation

- No
- 1 year
- 3 year

Select the product you want to purchase

Calculate costs

Reservation quantity (3 recommended) ⓘ

3 ✓

Cost of reservation (US... ⓘ

4,406

1,469 per reservation

Pay-as-you-go cost (U... ⓘ

960

Remaining

New estimated total (... ⓘ

5,367

Was 8,720

Savings (USD) ⓘ

3,353

Utilization ⓘ

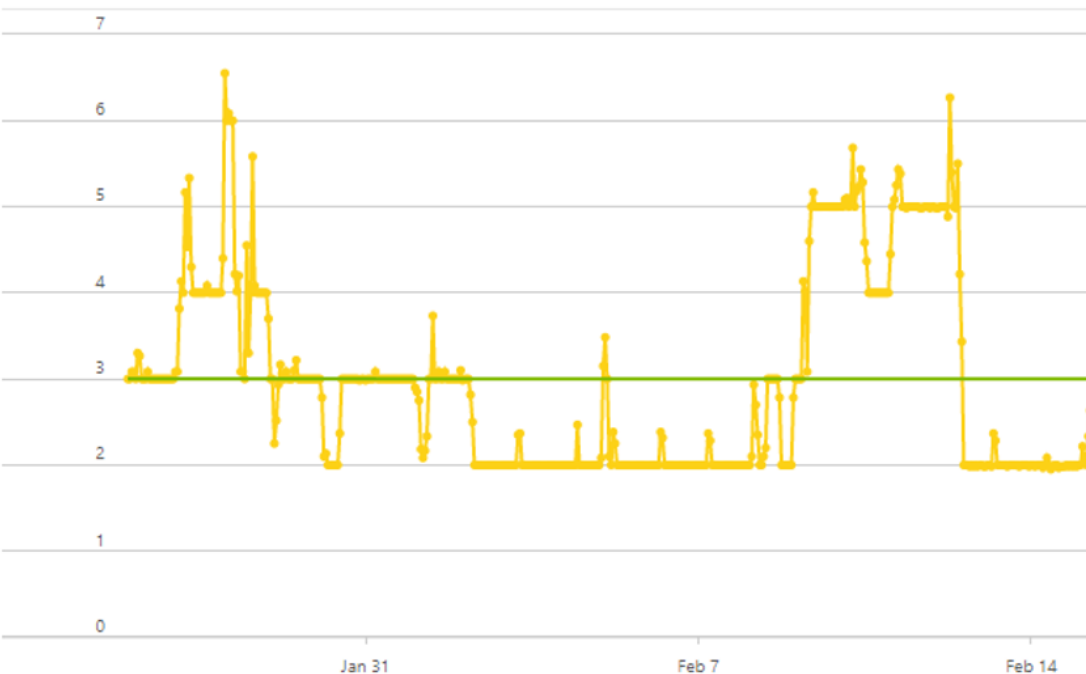
86%

Analyze data

Utilization over time

Savings by quantity

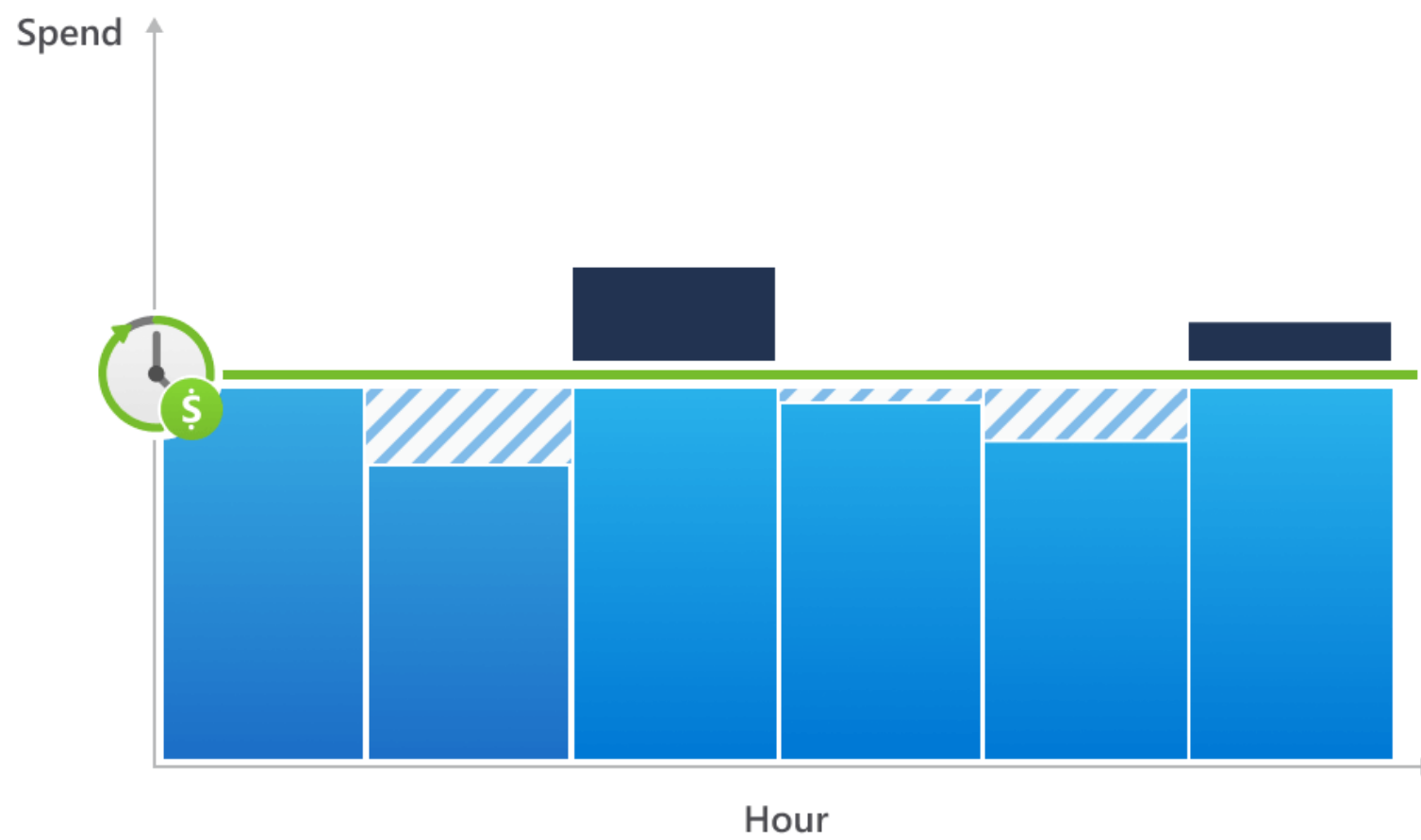
This chart shows the quantity (Y-axis) of Standard\_F2 in region Southeast Asia that was charged using on-demand pricing.



#AzConfDev



# Cost Toolset - (Re)use



■ Hourly amount ■ Savings plan prices ■ Pay-as-you-go-prices ▨ Unused savings plan

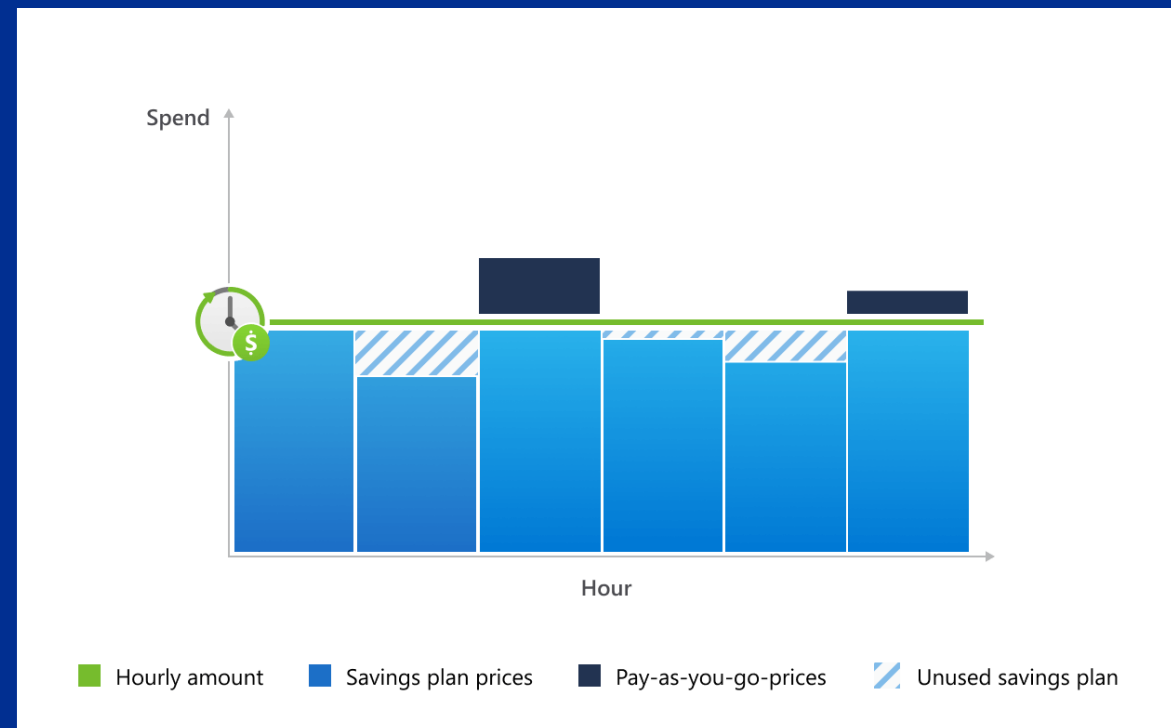
Microsoft  
Ignite

October 12-14, 2022

# Cost Toolset - (Re)use

## Azure saving plans

- Pay the full hourly amount, even if usage is less. Incur no additional costs—usage is covered by the plan.
- Usage up to your hourly amount is billed at lower prices and included in the cost of your plan. Additional usage is billed at pay-as-you-go prices and invoiced separately.



# Cost Toolset - (Re)use

Azure saving plans – how to start



1



Review your [savings plan recommendation](#) in the Azure portal based on recent usage.

2



[Purchase](#) a savings plan. Choose your hourly commitment, term length, billing frequency, and where to apply savings.

3



Start saving immediately. Your savings are automatically applied every hour you use select services.



#AzConfDev

# Demo

# Design for costs – Four pillars of economic cloud design



**Twan Koot**

Lead Performance Engineer & Codesmith  
bij Ordina



#AzConfDev



**Michiel Hamers**

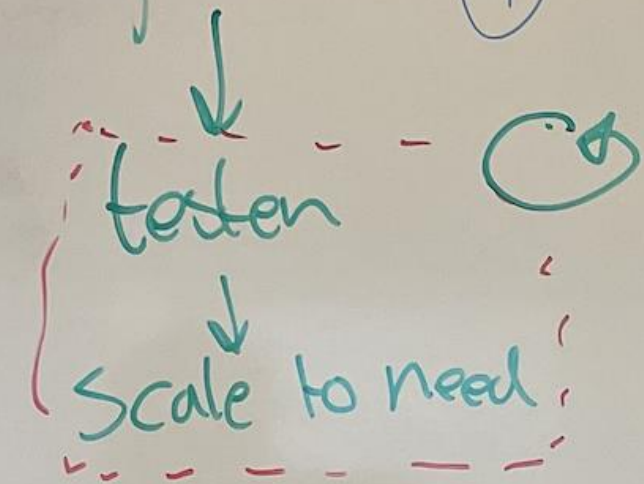
Solution Lead Microsoft (Azure)  
Cloud | Trainer | Speaker



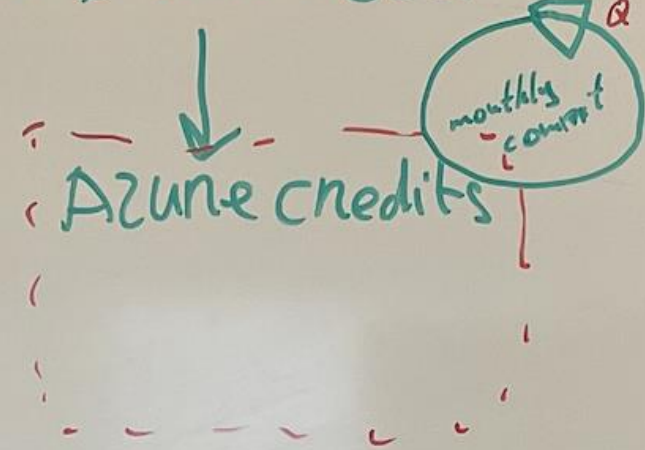


# whiteboard

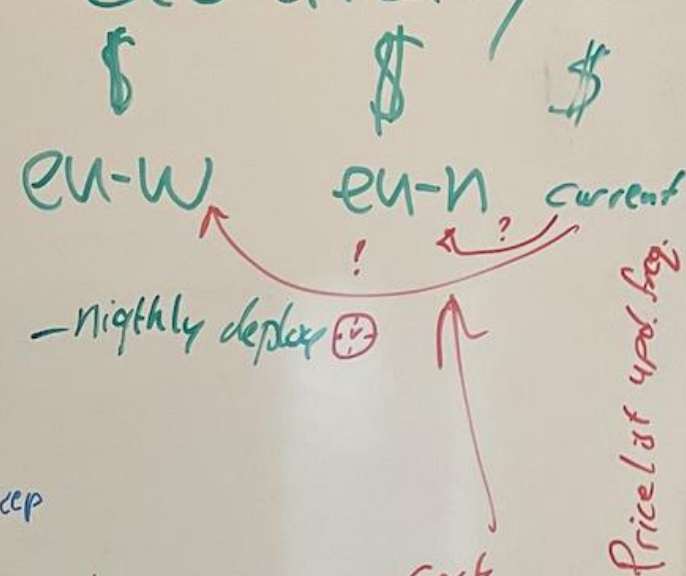
- PIVV ①



- Ø User cost ②



- Geo efficiency ③

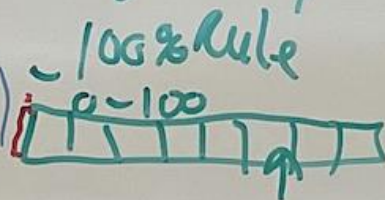


- Scaling out/in <sup>not up</sup> add down  
- runbook

- monitoring <sup>users focus</sup> ④

- Full backup vs. config <sup>ARM</sup>

- tag policy



Architect for costs <sup>low</sup> <sup>High impact</sup>  
 1. default service plan  
 2. Zero user costs  
 3. Geo Efficiency  
 4. User Focus

500 250  
 - cost/Quality  
 - automate all steps



# Geo efficiency - ms vs costs

```
ping d4pcentralindia.centralindia.cloudapp.azure.com -n 100
```

```
ping d4psouthindia.southindia.cloudapp.azure.com -n 100
```

```
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Reply from 4.247.26.204: bytes=32 time=3ms TTL=114
Ping statistics for 4.247.26.204:
    Packets: Sent = 100, Received = 100, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 2ms, Maximum = 4ms, Average = 2ms
```

```
Command Prompt
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=28ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=28ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=28ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=30ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=30ms TTL=115
Reply from 20.235.124.76: bytes=32 time=29ms TTL=115
Reply from 20.235.124.76: bytes=32 time=34ms TTL=115
Reply from 20.235.124.76: bytes=32 time=28ms TTL=115
Ping statistics for 20.235.124.76:
    Packets: Sent = 100, Received = 100, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 28ms, Maximum = 34ms, Average = 29ms
```



# Cost Toolset - (Re)use - Tags

Use Tags for all your Azure Resources (and Resources groups).

Via Tags you can create cross charging overviews.



#AzConfDev



# Cost Toolset - Improve

Follow architecture framework scalability principles

- Understand the challenges of distributed architectures
- Run performance testing in the scope of development
- Continuously monitor the application and the supporting infrastructure
- Identify improvement opportunities with resolution planning
- Invest in capacity planning



SMALL  
STEPS  
EVERY  
DAY



#AzConfDev



Get all resources of a specific subscription and datacenter  
Get all costs of previous full month  
Get all costs of all resources with specified datacenter