Design for costs

Four pillars of economic cloud design

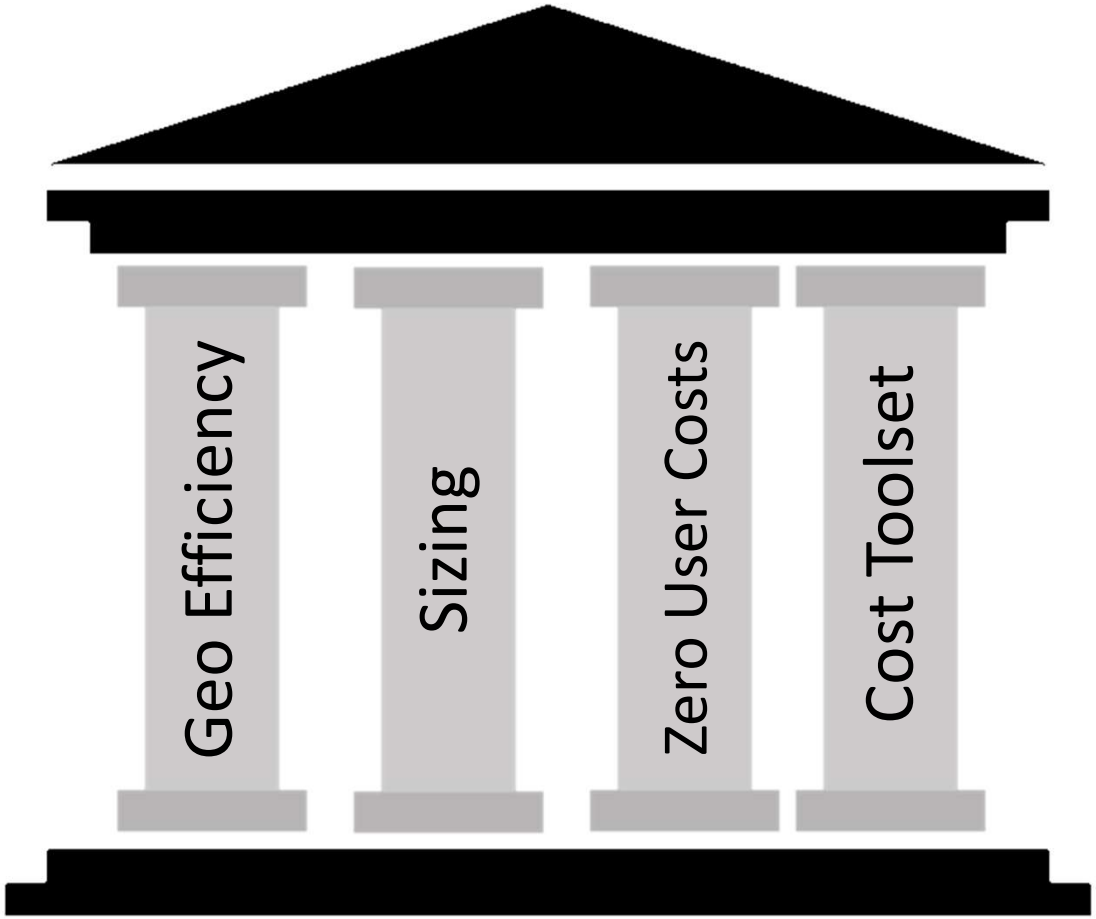
Twan Koot

Michiel Hamers

Agenda

VISUG

- Introduction
- Why?
- 4 pillars
- Tips, Tricks & QA



VISUG

Michiel Hamers

Solution Lead Microsoft (Azure) Cloud &
Codesmith - Ordina

VISUG

Twan Koot

Lead performance engineer & Codesmith - Ordina

Disclaimer

VISUG

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?

VISUG


I'VE FINALLY FOUND IT... AFTER 15 YEARS
THE SCROLL OF TRUTH!
Cloud isn't cheaper then on prem
NYEHHH
imgflip.com


How did you end up in poverty? Gamble? drugs?
I left an ec2 instance on.

Geo Efficiency

The first pillar

Geo Efficiency - Geo what ??

VISUG

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

Geo Efficiency - It's all about the money

VISUG

There is a big difference in datacentres pricing between each region.

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

West Europe: €824,92 North Europe: €764,28

Switzerland west: €1.179,68

Brazil SouthEast: €1.419,96 South Africa North: €909,82

South India: €965,91

Geo Efficiency - Impact at scale

VISUG

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

Geo Efficiency - Picking an Azure region

VISUG

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

Geo Efficiency - Let's compare

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

West Europe: €824,92 ~8% Difference North Europe: €764,28

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

Central India: €450,93

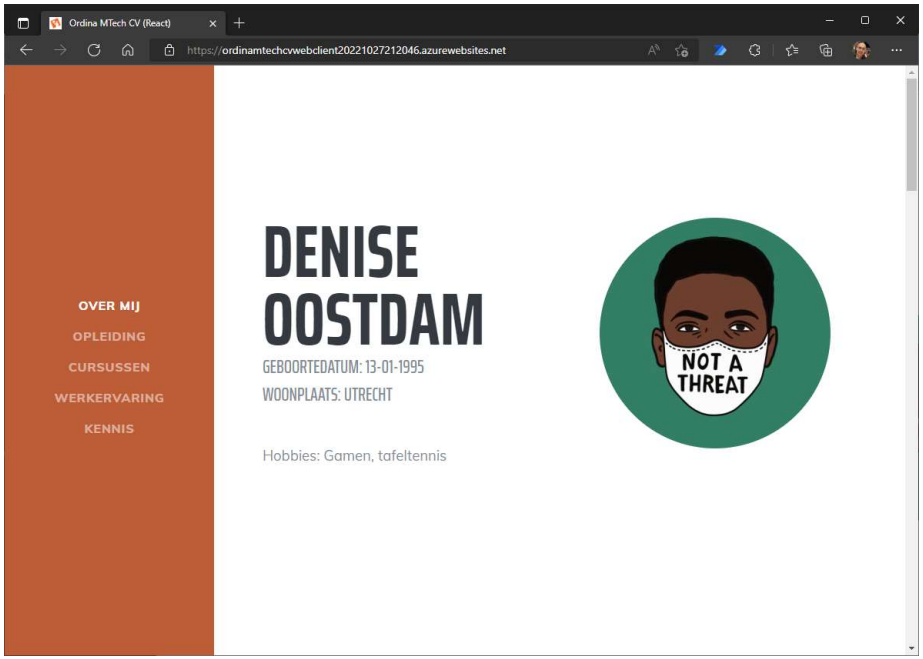
This means a potential save of 45%!!

Geo Efficiency - Latency impact

VISUG

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

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



Geo Efficiency - Latency impact

Let’s compare Azure South vs Central India

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

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

Geo Efficiency - Latency impact

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?

VISUG

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 recourses should follow the following these steps:

1.) Choose the region based on data residency/complaince

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

Sizing

The second pillar

Sizing

Tiers*

- Free
- Shared
- Basic
- Standard

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

Sizing VMs

VISUG

Type	Sizes	Description
General purpose	B, Dsv3, Dv3, Dasv4, Dav4, DSv2, Dv2, Av2, DC, DCv2, Dpdsv5, Dpldsv5, 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, Epdsv5, 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).

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***

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

Sizing - Naming

VISUG

[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 constrained vCPU capable size
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 Ddv4 and Ddsv4-series 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 Ultra SSD (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.

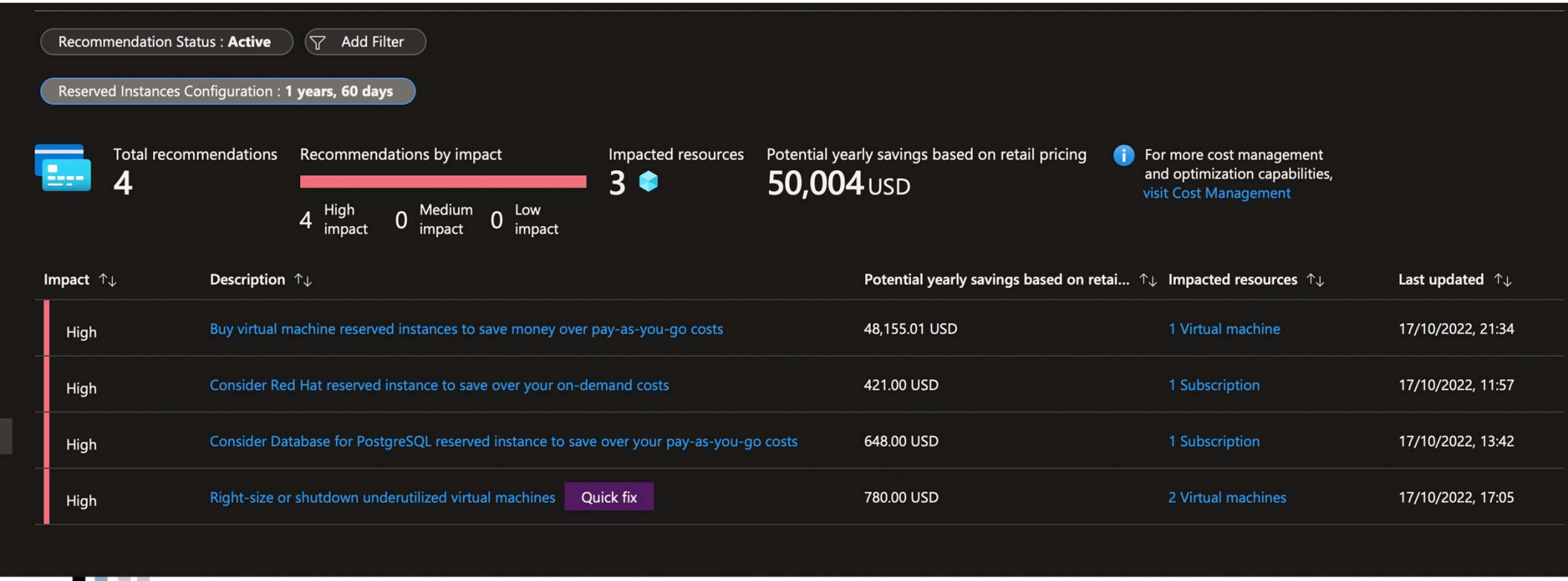
VM Name	vCPUs	Memory (GiB)	Linux Cost	Windows Cost
Standard_D2	2	7	135670.50	201711.26
Standard_D2_v2	2	7	121313.74	187354.50
Standard_D2_v3	2	8	96907.50	170126.21
Standard_D2_v4	2	8	96907.50	162948.26
Standard_D2_v5	2	8	96907.50	162948.26

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

Sizing - Azure advisor

VISUG

Azure advisor is for VMs



Sizing - Monitoring advisor

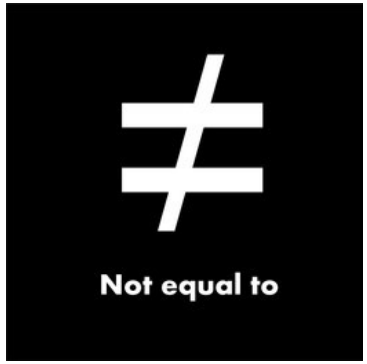
VISUG

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

Sizing - Prices

VISUG

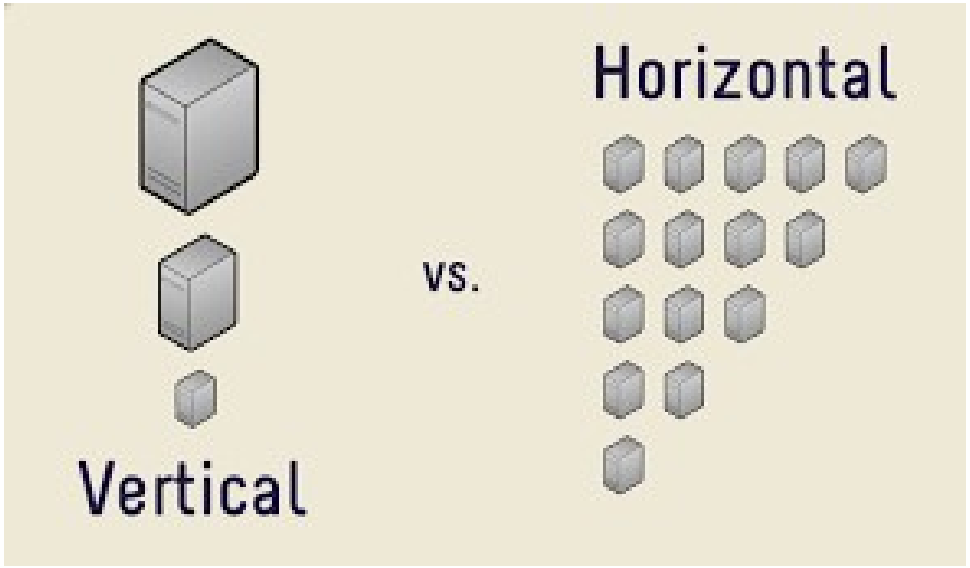
- Listed prices
- Personal prices
- Discounts



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)!



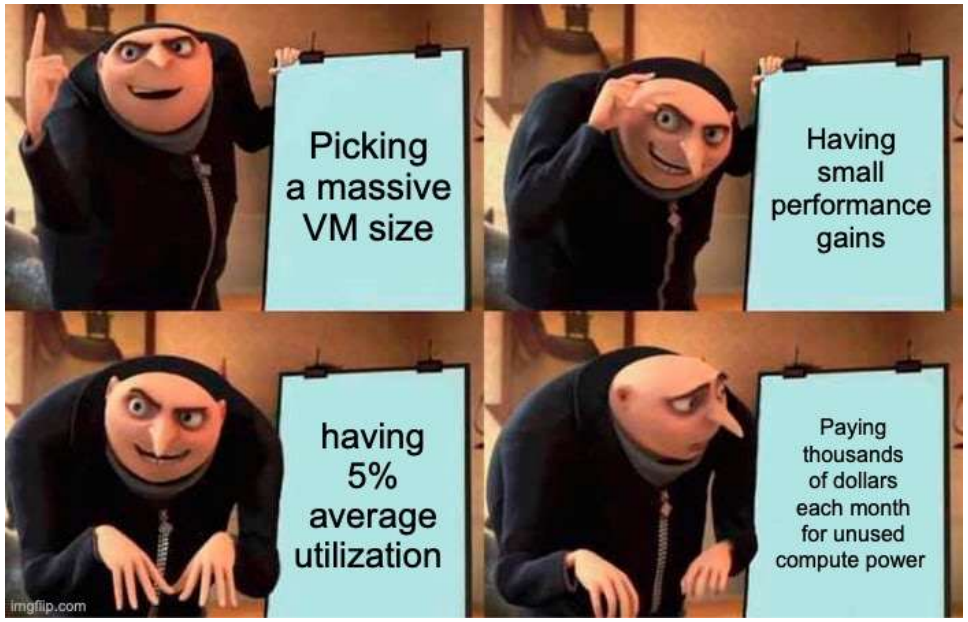
Sizing - Picking the right size

- 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”



Sizing - Picking the right size

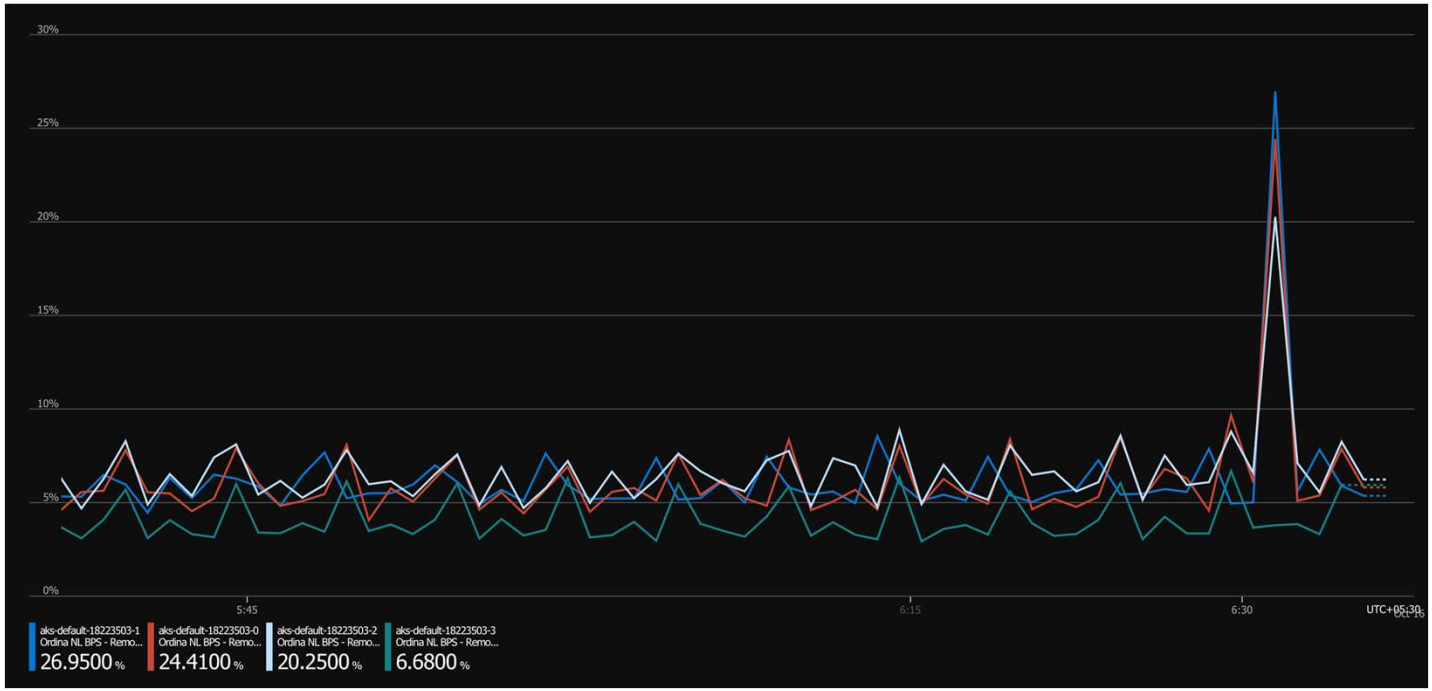
Look at response times instead of utilization

Running daily load between 60~75% utilization should be the goal

Sizing - Monitoring

VISUG

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



Zero User Costs

The third pillar

Zero User Costs - Let's Explain

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

Zero User Costs - Let's Explain

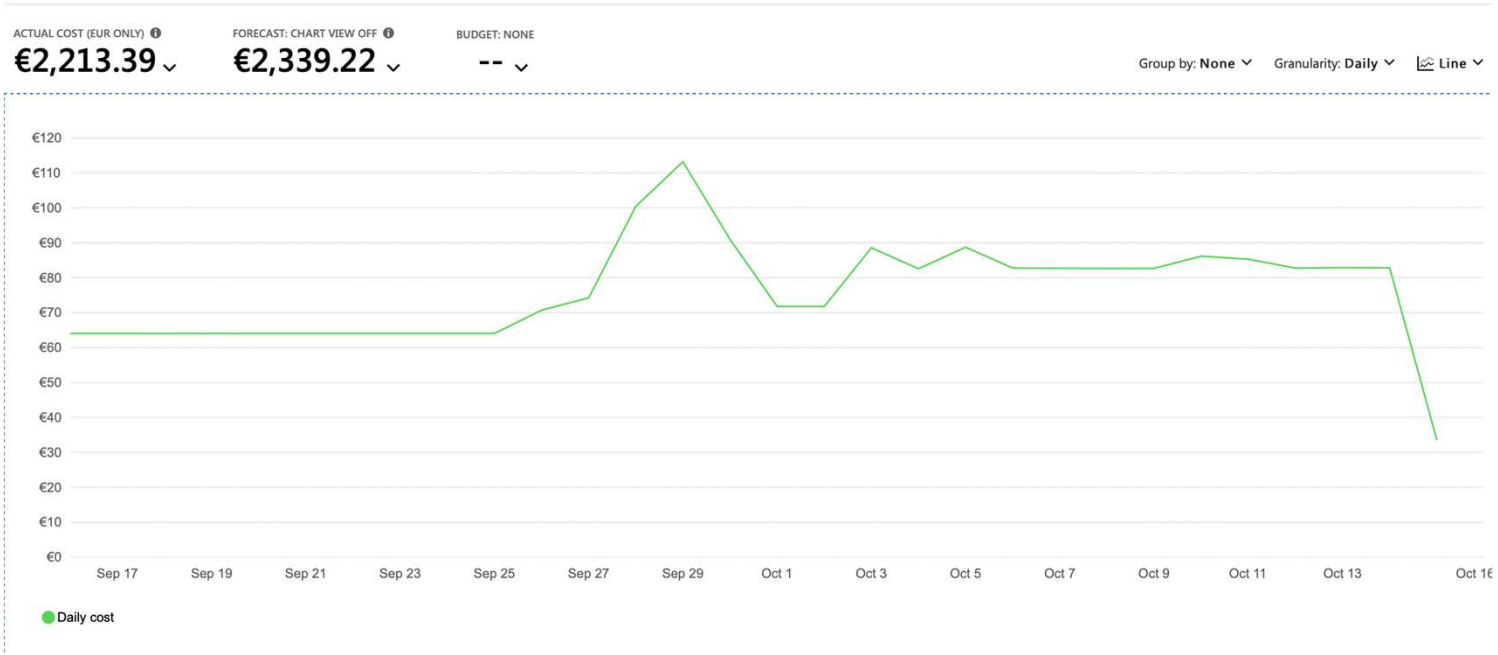
VISUG

- 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

VISUG

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



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

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 de- scribes the core of the solution to that problem, in such a wat that you can use this solution a million time over, without ever doing it the same way twice"
(Alexander et al., 1977).

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

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:

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

VISUG

Cost Toolset

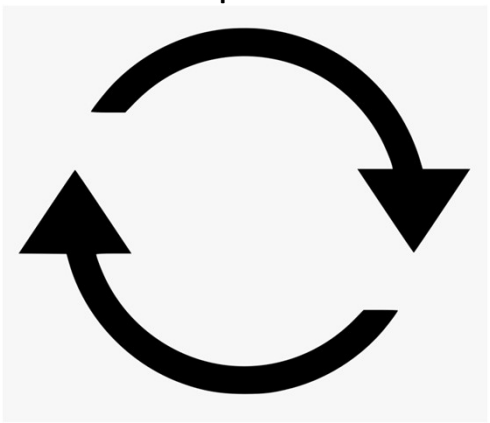
The fourth pillar

Cost Toolset

VISUG

Improve

Repeat



Reuse

Cost Toolset - (Re)use

When to use Hot/Warm/Cold data path?

Path	Requirement
Hot data path	• When data requirements are known to change frequently • When processing or displaying data in real time
Warm data path	• When you need to store or display a recent subset of data • Used for data that is consumed for small analytical and batch processing
Cold data path	• When data is rarely used. The data might be stored for compliance or legal reasons • Used for data that is consumed for long term analytics and batch processing

Cost Toolset - (Re)use

VISUG

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

Data storage prices pay-as-you-go

All prices are per GB per month.

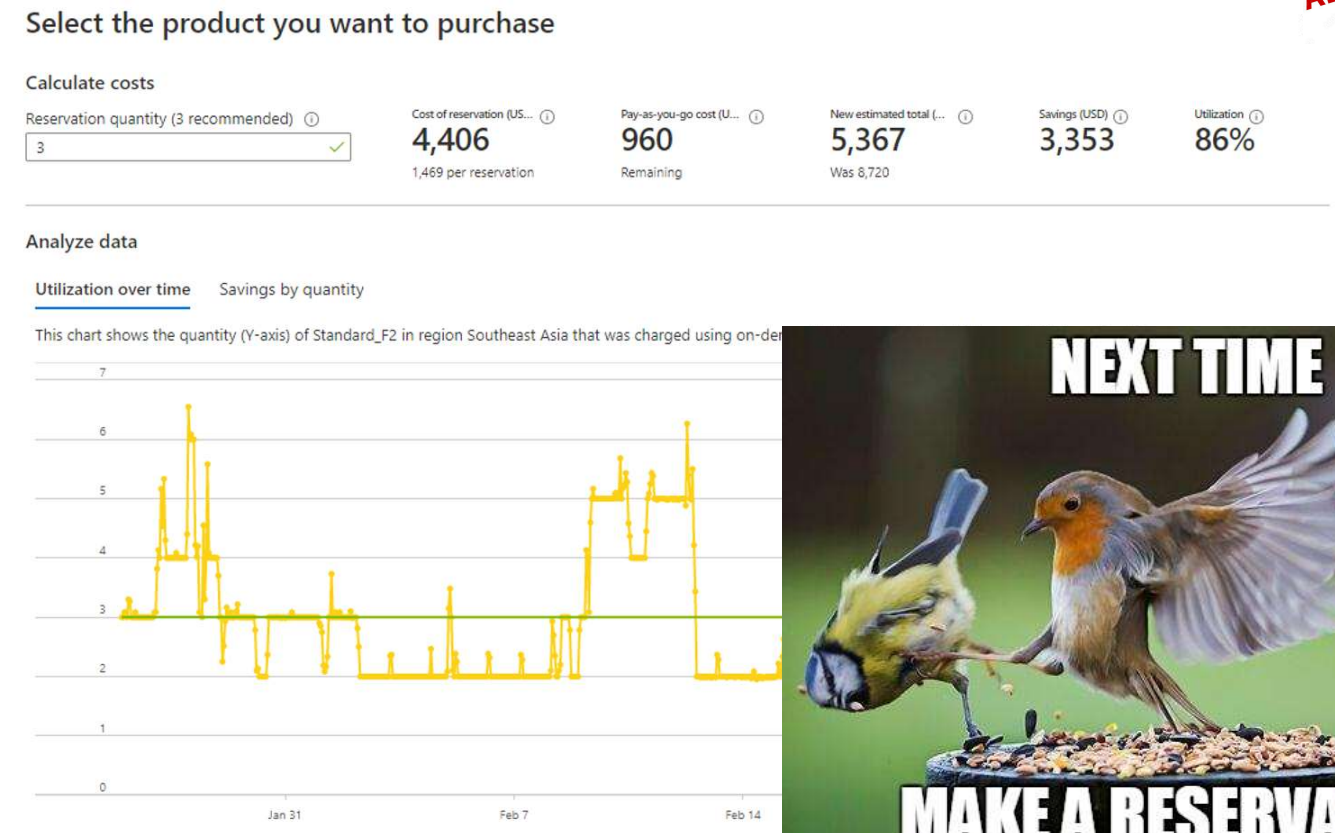
	Premium	Hot	Cool	Archive
First 50 terabyte (TB) / month	₹17.20833 per GB	₹1.6389 per GB	₹0.90139 per GB	₹0.16389 per GB
Next 450 TB / month	₹17.20833 per GB	₹1.5734 per GB	₹0.90139 per GB	₹0.16389 per GB
Over 500 TB / month	₹17.20833 per GB	₹1.5078 per GB	₹0.90139 per GB	₹0.16389 per GB

Cost Toolset - (Re)use

VISUG

Azure reserveration

- No
- 1 year
- 3 year



Cost Toolset - (Re)use

VISUG

Azure saving plans

Cost Toolset - (Re)use

VISUG

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

VISUG

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.

Demo

Thank you!

VISUG

codit

involved

Capgemini

Xpirit

team4talent

ORDINA

AllPhi

Enabling the next generation software developers

AXXES

Microsoft

Thank you!

VISUG

Design for costs – Four pillars of economic cloud design

Twan Koot
Lead Performance Engineer & Codesmith bij Ordina

Michiel Hamers
Solution Lead Microsoft (Azure) Cloud | Trainer | Speaker