

**What to look out for when  
migrating your data to the cloud:  
is it *REALLY* the right choice? 🧐**

# Sponsor & Org



UNIVERSITÀ DEGLI STUDI DI PARMA



# Emanuele Meazzo

- Old time speaker
- Manager Data Science & Intelligence
- Acer EMEA

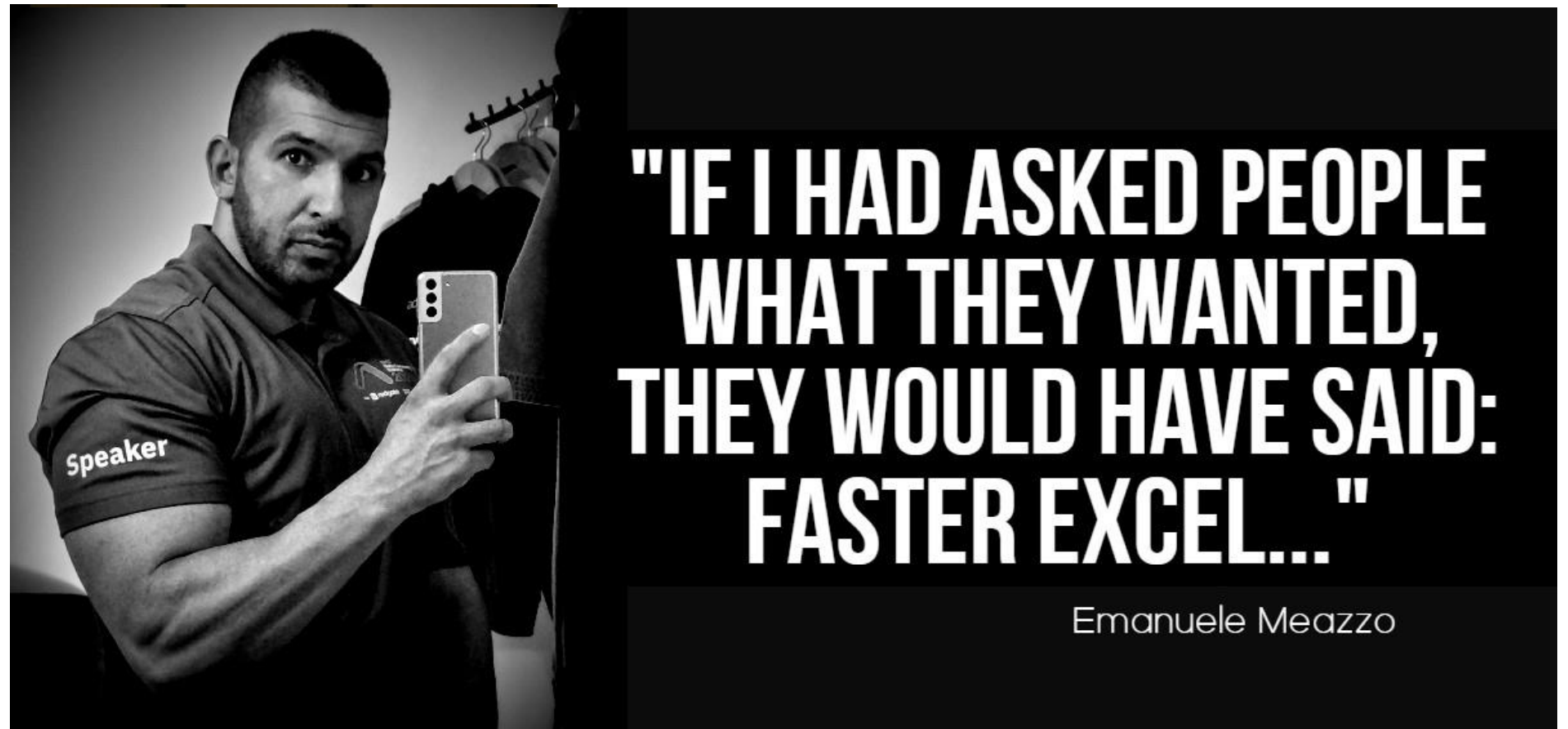
@619619



[linkedin.com/in/emanuelemeazzo](https://www.linkedin.com/in/emanuelemeazzo)



[tsql.tech](https://tsql.tech)



# What's this about



This is a not deeply technical presentation



But it's cool since you can bring some of these points to your boss



Hopefully, everything should be trivial for you



If you're already "in the cloud" and any of this isn't trivial, sit down, we have to talk



The argument is so vast we could talk for hours, but there's only 60 minutes, there will be gaps, we can continue at a different time, just ping me



Unfortunately, you're going to leave this session with more questions than answers

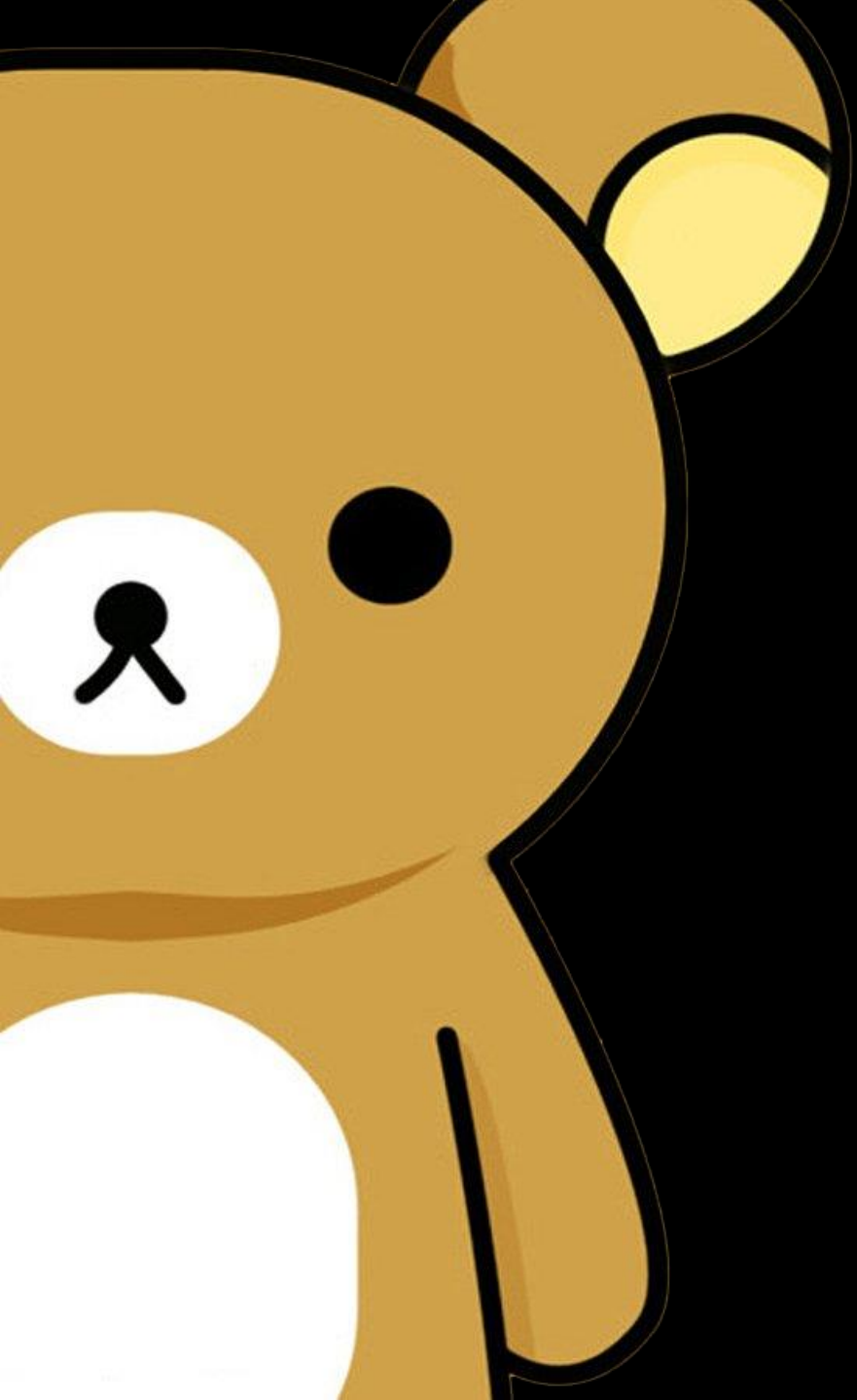
Slides are text rich in order to be of reference for the future, you don't need to read it all



# **FIRST REACTION SHOCK**

“If someone asks me what cloud computing is, I try not to get bogged down with definitions. I tell them that, simply put, cloud computing is a better way to run your business

—Marc Benioff,  
Founder, Salesforce

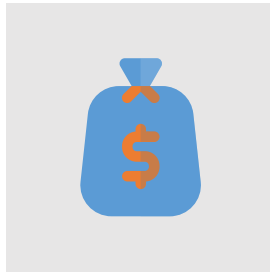


**But first, *bear* with  
me for a second**

**We need a refresher  
on a couple of  
non-technical aspects**

# CapEx vs OpEX

AK Why the business wants to move to cloud



## CapEX

Capital Expenses

**Consist of the funds that companies use to purchase major physical goods or services that the company will use for more than one year**



## OpEX

Operational Expenses

**Represent the day-to-day expenses necessary to run a business.**

# Types of Cloud Services

## IaaS

Infrastructure  
as  
a  
Service

Getting Rid of  
Physical  
Hardware

## PaaS

Platform  
as  
a  
Service

Getting Rid of  
installations  
and updates

## SaaS

Software  
as  
a  
Service

Getting rid of  
anything but  
the software  
configuration



# New companies and the Cloud

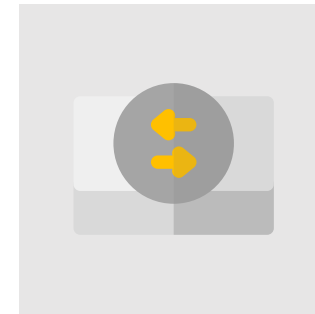
It's usually a no-brainer to start cloud-first

## Cloud as a startup advantage



- No upfront investment in architecture
- Lower capital risks
- Faster creation of a Minimum Viable Product
- Faster business pivoting

## Cloud as a business model



- A SaaS product is easier to try and to adopt
- SaaS Profit = EBITDA

# Existing companies and the Cloud

This is where the real planning and careful evaluation is needed  
The following questions must be asked first:

1

What is my goal?  
Cost savings?  
New Product?  
Faster  
Performances?

2

What does my current  
infrastructure look like?  
Is the hardware at EOL?  
Or is it just slow?

3

What technologies my  
team knows deeply?  
Am I willing to invest in  
their formation and/or  
in new hires?

# The State of the Cloud

- 81%** Of enterprises indicate that managing the cloud spend is a challenge
- 60%** Of organizations use cost efficiency and savings to measure cloud progress
- 64%** Plan to optimize the existing use of cloud (cost savings), making it the top initiative for the fifth year in a row
- 59%** More than 59 percent of enterprise workloads and data are or are expected to be in a public cloud within 12 months
- 28%** It's estimated that organizations waste 28% of cloud spend, are over budget by an average of 18% and expect cloud spend to increase by 29% next year

# Speed

# 50+

Azure updates per Quarter



**Kalm**

Most of the changes don't require any action, the cloud vendor innovates for you

## The Cloud is **FAST**



**Panik**

All these changes are difficult to follow

# Multiple Vendors?

## Everything in the same Cloud?

### PROS

- Lower TCO (possibly)
  - Bundle licenses & discount tiers
  - No egress charges if in the same geographic area
  - Reduced data movement
  - Reduced ETL complexity
- Easier to focus on optimizing a specific platform
- Ability to hire specialized developers

### CONS

- Suboptimal feature set (possibly)
- If your provider decides to deprecate a critical feature for your business, you're screwed
- In case of a cloud provider that's no longer competitive, the whole IT stack needs to be moved over

## Use multiple Cloud Vendors?

### PROS

- Lowered Risk (?)
- Higher Flexibility
- Best of breeds solutions

### CONS

- Higher Costs
  - Egress fees
  - ETL costs to integrate data
  - Knowledge
- Security (SSO)
- Governance
- Visibility
- Monitoring and Logging

**Having said this: 92% of Enterprises have a multi-cloud strategy**

# Trivia



**Things you can say to your cloud vendor but not your partner in bed**



**I want it to be FAST and CHEAP**

**Ok, let's go a bit deeper  
into the topic**

**Finally.**

# The DataLake trap (aka continuous upselling)

I've seen countless companies burning large amounts of cash in search for the “big data” solution, only because the management\* was driven by the Big Data hype

# 60-85%

[Source](#)

Is the reported failure rate of Big Data enterprise projects

Avoid dumping data into a big bucket just because “now you can” using cloud tools.  
Saving money means first and foremost not throwing them away in a project that will never see the light of the day.

Ask yourself: What is the problem that I'm trying to solve with this?



# The trap, continued

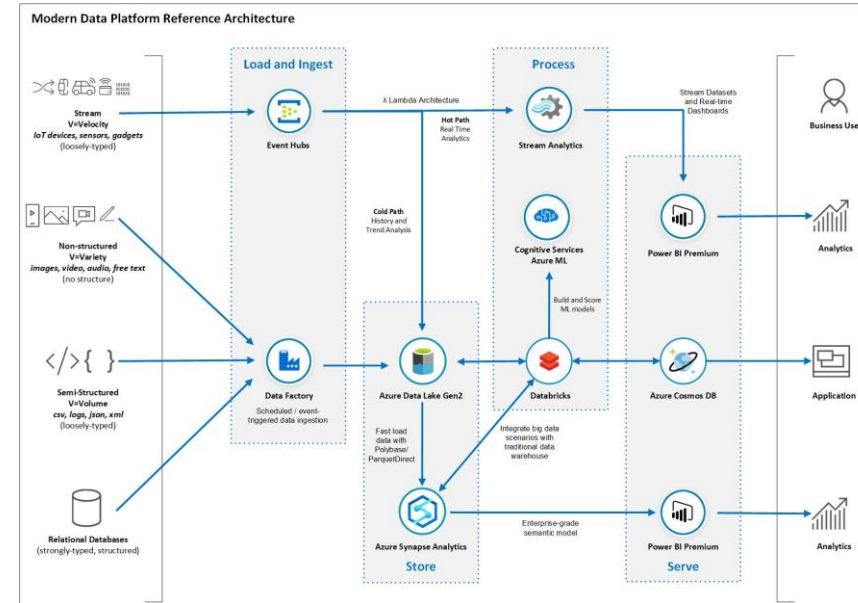
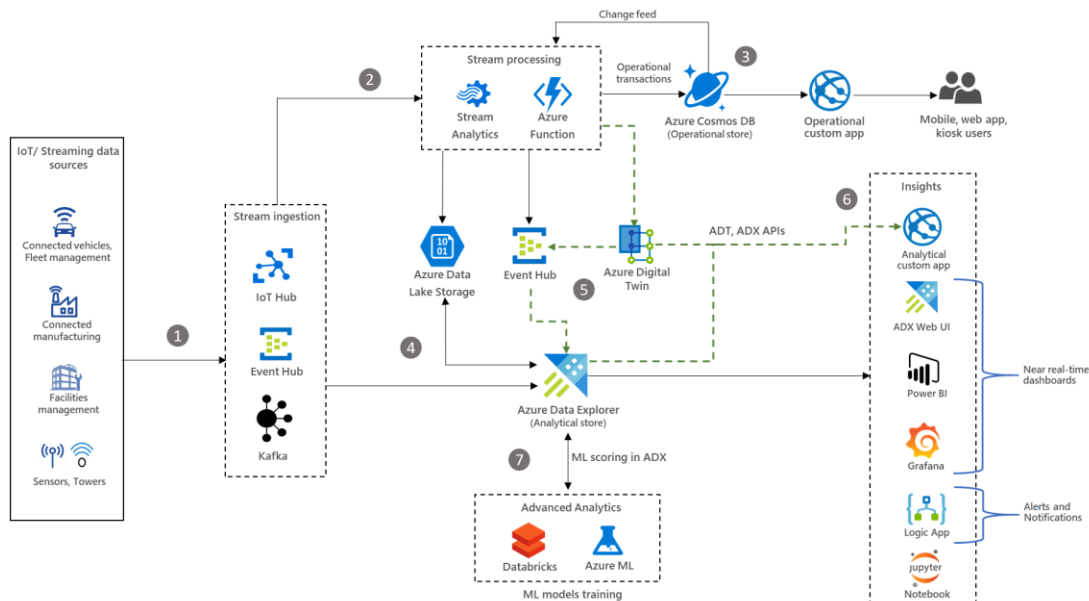
Choose the right technology for you.

The market is full of buzz-words, everything is real time, petabyte scale, streaming rich data.

The first question should be, what do I need to do?

What's the problem I'm trying to solve?

What's the business requirement behind all of this?



In the cloud architecture examples you see on the web, you'll see lots of services communicating with each other

Of course, any vendor would be very happy to sell you all the cloud services you need, or even better everything they think you need

**Sorry,  
probably,  
your apps need to change**

# My apps need to change !?!?



www.lunarbaboon.com

That's the main point of all of this, there is no real Cloud transformation without a change in your ways

No more “It always has been done like this”  
If you want to do what you’ve always done, but in the cloud, you’ll rarely see real benefits from it.

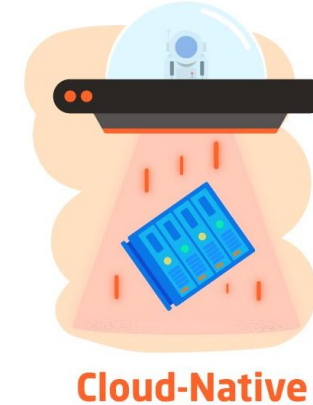
Now, let’s see some of the typical hurdles of cloud adoption ☹☹

# YES.

# Thinking why you're Lifting & Shiftin'



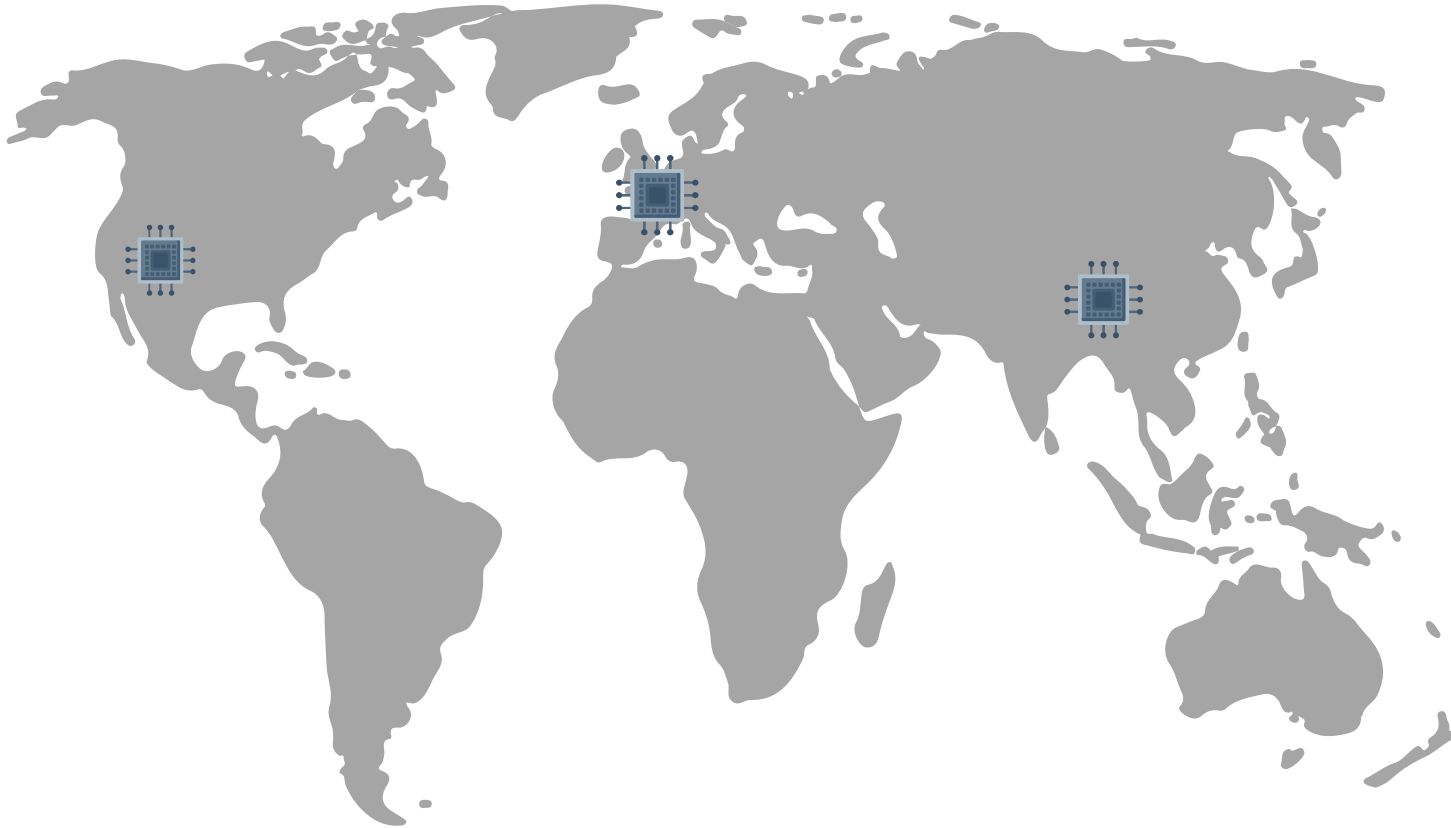
Vs



- A classic business *move* to get to the cloud “fast”
- Can be a good **short-term** solution (or a long term one for non business critical systems)
- Good move for legacy system with **EOL hardware** with already an hardware **refresh** planned
- Infrastructure costs are higher than continuing to use your servers, if you already own them
- Cannot use 90% of the features that makes the cloud more efficient
- You're basically renting **someone's else server** instead of your own (also, you're renting their network)

# Mind the Network and Expect Latency

Especially if you Lift&Shift parts of your application/data to the cloud, latency can bite you



Not necessarily because cloud latency is bad

Maybe you're used to the sub-ms latency of your server which was sitting in the closet behind you, and now your chatty application is trying to get data from the other side of the world.

For existing apps, verify the level of chattiness and how latency affects them

# Mind the Egress Fees

When you start moving to the cloud, you'll have flows of data in and out of cloud resources  
Be careful, it's not like your good 'ol datacenters, you're going to pay for it

While importing data is mostly free, moving data from the cloud to on prem always costs you money

Azure Sample Pricing:

- 100GB-10TB/Month: \$0.08 per GB ~ \$800/month (10TB)
- 10TB/40TB/Month: \$0.065 per GB ~ \$3466/month (50TB)

So, **be careful when moving big amounts of data.**

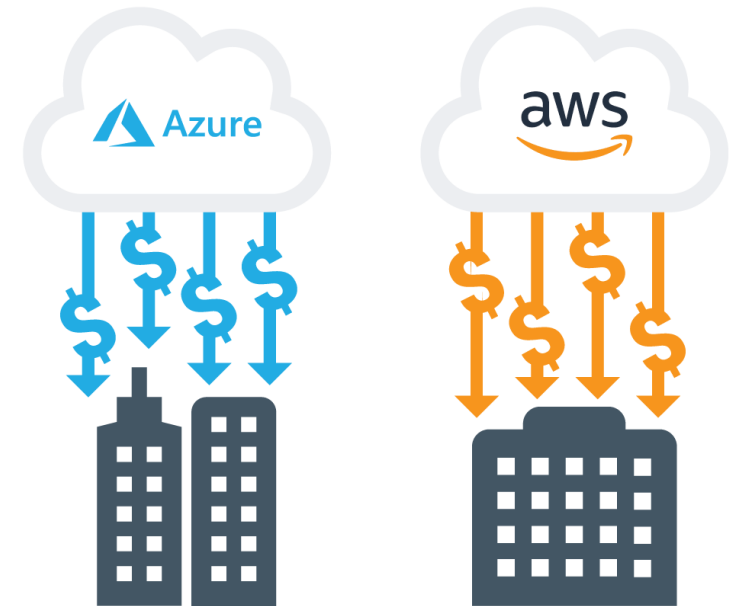
But wait, there's more!

**You pay for intra-cloud data movement too!**

And I'm not talking like between different cloud providers

Inter-region data transfer is billed (albeit at a lower rate, if you're in US or EU)

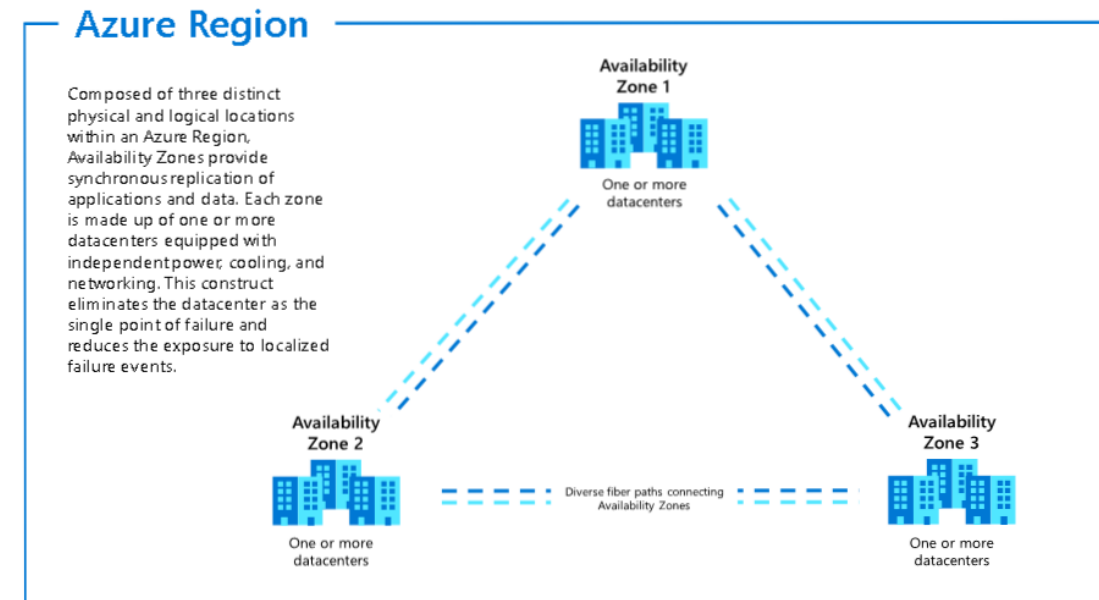
Finally, data transfer between Availability Zones is billed



# Control where your data is

After the two previous slides, you should have realized that data location is somewhat important  
It's all a game of balancing the following:

- **Region exclusivity:** Keep all traffic within the same region.  
If traffic needs to exit a region check and choose the region with the lowest transfer rates
- **Region Proximity:** The closer the cloud region to your premise and/or users, lower the latency
- **AZ exclusivity:** In both AWS and Azure traffic within the same Availability Zone is free.
- **Geo Replication:** Critical data, that needs a very tight SLA should be replicated, make sure it's worth it



# Consider Scaling Services

You can get most of the cost savings by simply being able to scale up/down cloud resources, or even better being able to **turn them off** completely

If you're using a cloud VM like an on-prem one, in a never-changing configuration, what's the point of it being physically more distant from your premise and managed by a third party?

At the same time, the scaling considerations to make are Different than on-prem:

- How can this service scale?
- Can it be **turned off**?
- **How long** does scaling take?
- Does my application support **horizontal data scaling** or does it need to be baked into the data layer?
- Does this dev instance really need to stay on 24/7 ?





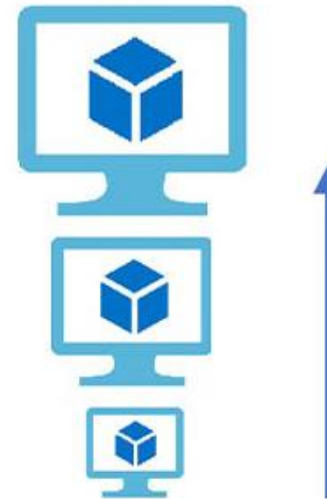
# More considerations on Scaling

Like any other solution, however, scaling introduces some problems to solve, for example:

- Depending on the service, **scaling up and/or down may not be seamless**, and **service interruptions** may happen (from a simple drop of in-flight connections to various minutes of service interruption)
- Due to the above, any cloud-ready application should integrate a **retry logic** for any call to any service, especially to the persistence layer
- Does the service **scales automatically** up and down depending on the load, or you have to **programmatically scale** it yourself?
- While vertical scaling can be almost transparent (except the downtime), a horizontal scaling isn't, if your application isn't already in the cloud mood; do you even know **how it scales**?
- If your data starts getting spread across various locations, you really have to think about how to **distribute** it

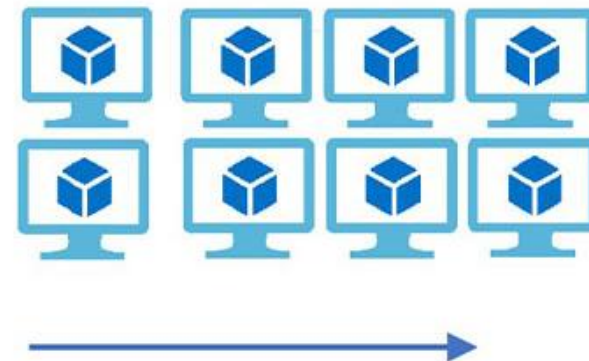
## Vertical Scaling

( Increase size of instance (RAM , CPU etc.) )



## Horizontal Scaling

( Add more instances )



# Performance is (literally) money

Scaling up usually happen for two reasons:

- A good one: an **increased request volume** to your system
- A bad one: a **performance problem** that's being fixed by blindly throwing hardware at it



Since in the cloud we're basically renting someone's else hardware, using way more than necessary because no effort has been spent in optimizing the data access is an **expensive** mistake.

# Please 🙏 mind the IOPS/Throughput limits

- When selecting a Paas database, read carefully the whole spec sheet
- IOPS can become a limiting factor if not considered, when coming from OnPrem
- Of course these numbers are often buried inside the documentation of the product, but they need to be taken into consideration

Feature	General Purpose			Business Critical		
---------	-----------------	--	--	-------------------	--	--

## File IO characteristics in General Purpose tier

In the General Purpose service tier, every database file gets dedicated IOPS and throughput that depend on the file size. Larger files get more IOPS and throughput. IO characteristics of database files are shown in the following table:

File size	>=0 and <=128 GiB	>128 and <=512 GiB	>0.5 and <=1 TiB	>1 and <=2 TiB	>2 and <=4 TiB	>4 and <=8 TiB
IOPS per file	500	2300	5000	7500	7500	12,500
Throughput per file	100 MiB/s	150 MiB/s	200 MiB/s	250 MiB/s	250 MiB/s	250 MiB/s

# Reservations on Reservations

Reserving capacity on the cloud is a nice way to **save money**. All the major Cloud vendors offer discounts when opting for reserving a service for 1+ years, even up to 60% on the overall TCO

This is of course a great way to save money, but only if you do the following:

- Make sure that the resource **must** stay online for a percentage of time greater than the discount applied
- **Measure** your current workload
- Make sure that you're buying a resource be appropriate for the **whole reservation period** (While you can often exchange your reservation for another resource, it's usually only for upselling, not down)
- Understand that **you're reserving the price**, not the hardware itself



# Kill unused resources

Creating resource on the cloud is very easy, so much that's not uncommon to create something just for testing purposes and leaving it be for the time being.



The fact that some vendors makes it purposefully difficult to have an overview of everything that's deployed on a specific subscription surely helps with the above.

- **Review your cloud spending** every month
- Keep an eye on your **resource inventory**
- Use the integrated vendor tools for **budgeting and cost allocation** to identify which project/area is responsible for the most expenses and optimize it

# Storage Tiering & cleaning

**Storage is “cheap”** but when you start storing all kind of data in the cloud the bill starts to grow too

- If you have data that you **really** know you’re not going to use, get rid of it
- For all the rest, choose an appropriate storage tier:

“**Hot**” data is accessed frequently and/or needs a fast retrieval, hence it needs a fast storage

“**Cool**” data doesn’t have the same performance requirements, so, it can be stored someplace where you can wait a little bit to fetch your data

“**Archive**” is the cloud equivalent of tapes, you may wait for hours to get your data, but it’s cheap

PS: All your data needs to be archived as is or it can be aggregated first?



# The Cloud DBA

**Basically, still  
“It depends”**

**But now in Cloud,  
with more variables**

- Install, config SQL
- Install, config OS

- Performance Optimization
- Configure Scaling

role in the cloud

# Keeping Up



**Learning** has changed in the cloud days too

The technology changes so fast that's often not cost-effective for third parties to create in-depth courses about the newest cloud tech, as it could be obsolete by the time the course is ready

Big Books are a thing of the past for the same reason

To stay up to date, use these resources:

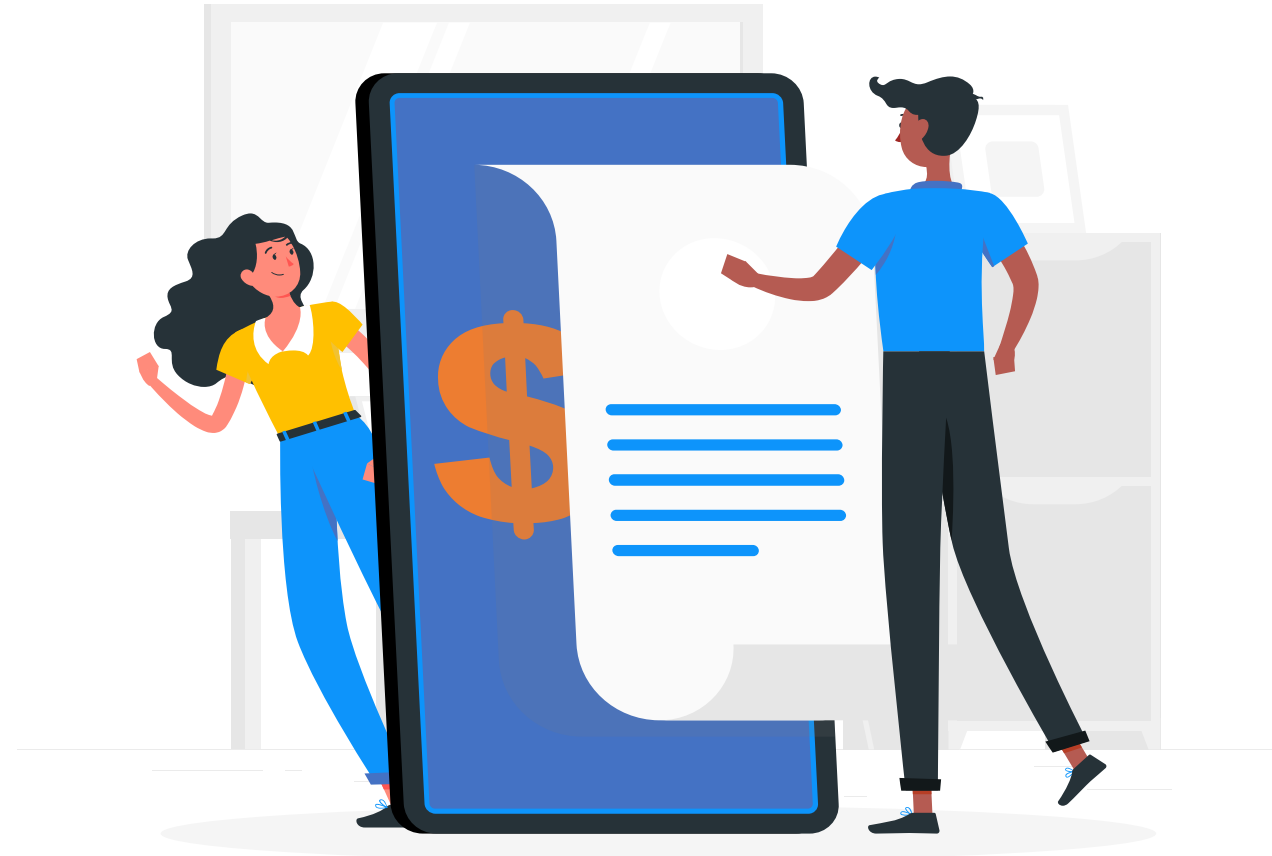
- **Blogs** (remember to check out mine!)
- Vendor Blogs & **Documentation**
- **Conferences** (like this one 😊)
- Vendor-organized “hands on” **events**
- Set up **your own** (small) instance and mess with it

Most importantly,  
learning needs to be **continuous**



# Azure Specific Tips

- Use **DEV/Test pricing** for, well, your dev/test resources (but consider Visual Studio Subscriptions cost)
- If you have an intermittent workload, take notice of the **Burstable VM types** (B-Series)
- Be sure to also **delete attached resources** when deleting a VM (this is now easier than ever thanks to new [delete options](#) gone GA this year)
- Microsoft is supposed to **Price-match** any other Cloud vendor for the same service every 3 months, if they haven't done so, try and ask why 😊





*I am inevitable.*



#37 PARMA 2023

**Grazie!!!**



