

A unique opportunity for you to be mentored by Amazonians



Batch 04

Week 3

22-July-2023



# Week 04 – 29-July



Marco Tamassia



Qusay Al-Maatouk

```
graph TD; A[Availability] --> B[Durability]; B --> C[Scalability]; C --> D[Reliability]; D --> E[Observability]; E --> A
```

**Availability**

**Durability**

**Scalability**

**Reliability**

**Observability**

Fundamentals  
Track

# Riding your bicycle



# A bicycle



# A spare wheel



# Downtime



Don't try this at home



## Availability



## Opening Hours

# COMPANY NAME

+420 123 123 12

6 7 8 9 10 11 12 13 14 15 16 17 18 19

**MONDAY**



**TUESDAY**



**WEDNESDAY**



**THURSDAY**



**FRIDAY**



**SATURDAY**



**SUNDAY**



## Opening Hours

# COMPANY NAME

+420 123 123 12

6 7 8 9 10 11 12 13 14 15 16 17 18 19

**MONDAY**

**TUESDAY**

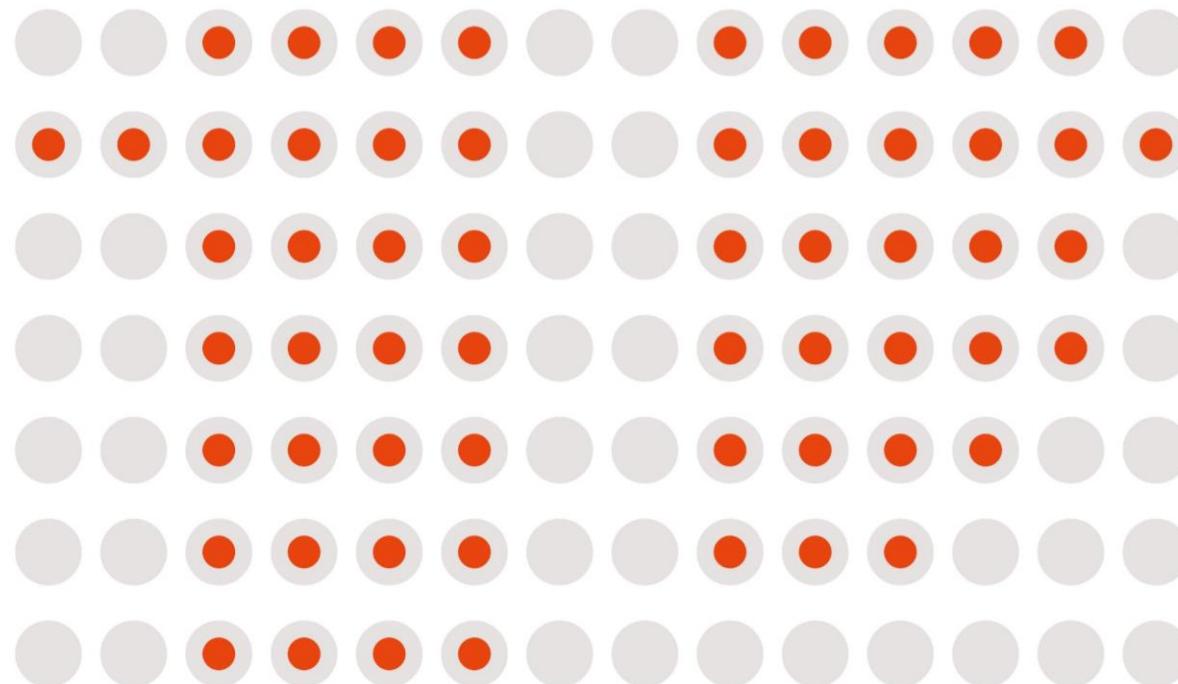
**WEDNESDAY**

**THURSDAY**

**FRIDAY**

**SATURDAY**

**SUNDAY**



[www.example.com](http://www.example.com) | [email@example.com](mailto:email@example.com)

# What is availability?

- Dictionary Definition
  - The quality of being available
- AWS Definition
  - Availability is the percentage of time that a workload is available for use.
  - Availability is reduced anytime that the application isn't operating normally, including both scheduled and unscheduled interruptions.
  - Availability is a percentage uptime (such as 99.9%) over a period of time (commonly a month or year)

$$\text{Availability} = \frac{\text{Available for Use Time}}{\text{Total Time}}$$

# How is availability calculated?

Availability	Maximum Unavailability (per year)	Application Categories
<u>99%</u>	3 days 15 hours	Batch processing, data extraction, transfer, and load jobs
<u>99.9%</u>	8 hours 45 minutes	Internal tools like knowledge management, project tracking
<u>99.95%</u>	4 hours 22 minutes	Online commerce, point of sale
<u>99.99%</u>	52 minutes	Video delivery, broadcast workloads
<u>99.999%</u>	5 minutes	ATM transactions, telecommunications workloads

Fault tolerance = big brother of availability



## Aircraft – Two Engines



# Two pilots



# Two navigation systems



# Different meals for pilots (served at different time)

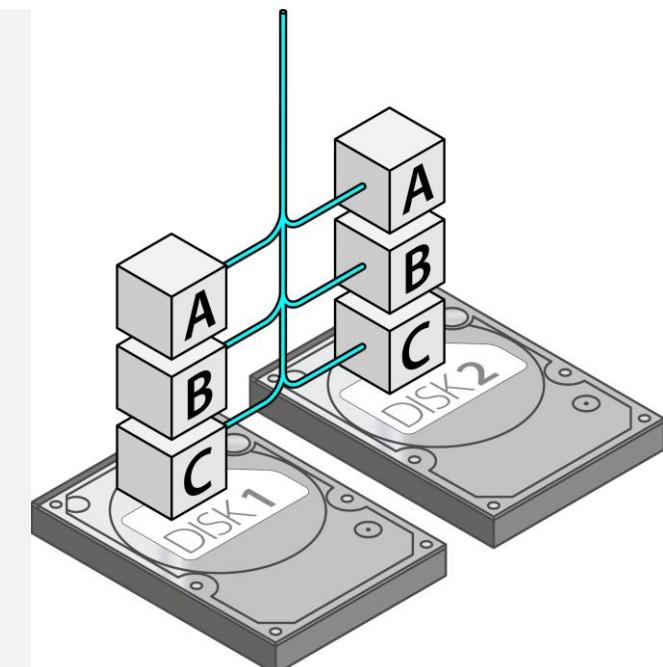
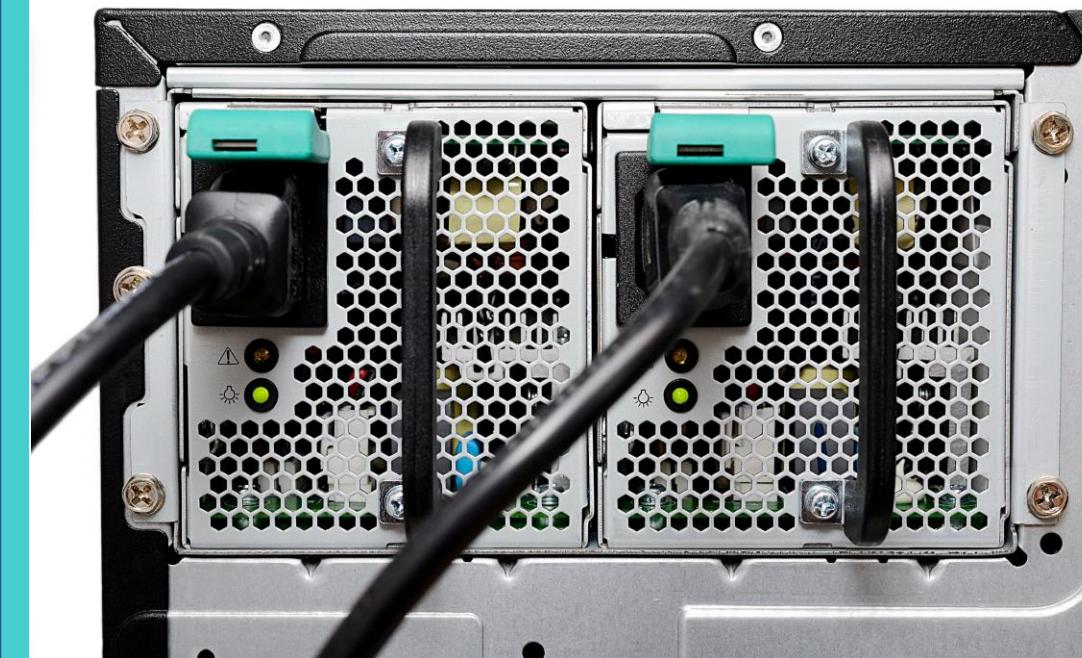


# Rule of two in cockpit



# Fault Tolerance

- Fault tolerance means that a system will almost always maintain uptime — and users will not notice anything different during a primary system outage.
- To implement fault tolerance, workloads use spare (or redundant) subsystems. When one of the subsystems in a redundant set fails, another picks up its work, typically almost seamlessly.



# Durability



## Durability

- A system that is durable is able to perform its responsibilities over time, even when unexpected events may occur. For example, a durable storage system will reliably store data without data loss.
- Amazon S3 Durability = Retention
- Amazon S3 has a data durability of 99.99999999% (11 nines) and 99.99% availability of objects over a given year.

# Amazon S3 Storage Tiers

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
<b>Designed for durability</b>	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)				
<b>Designed for availability</b>	99.99%	99.9%	99.9%	99.5%	99.9%	99.99%	99.99%
<b>Availability SLA</b>	99.9%	99%	99%	99%	99%	99.9%	99.9%
<b>Availability Zones</b>	≥3	≥3	≥3	1	≥3	≥3	≥3

**SCALABILITY**



## Excerpt from my book



Alex: I have a very funny example to share with you. Don't laugh. I am sure you all know about a fictional comic and movie character *The Hulk*. This character is most of the time seen as a normal human being (Dr. Bruce Banner), but when he gets emotionally stressed, he transforms into a muscular humanoid with immense physical strength. This is, I would say, a scale-up operation. And on the other hand, there is another character, called Ant-Man, who has a special suit with size-shifting powers, which can make him shrink to the size of an ant or allow him to be a giant. I would say that is elasticity. And yes, please don't judge me, I read lots of comics.



### Chapter 9 – I scale to the moon and back

Scaling – several resources, several types

Page 175

Link - <https://www.analogiescloud.com/>



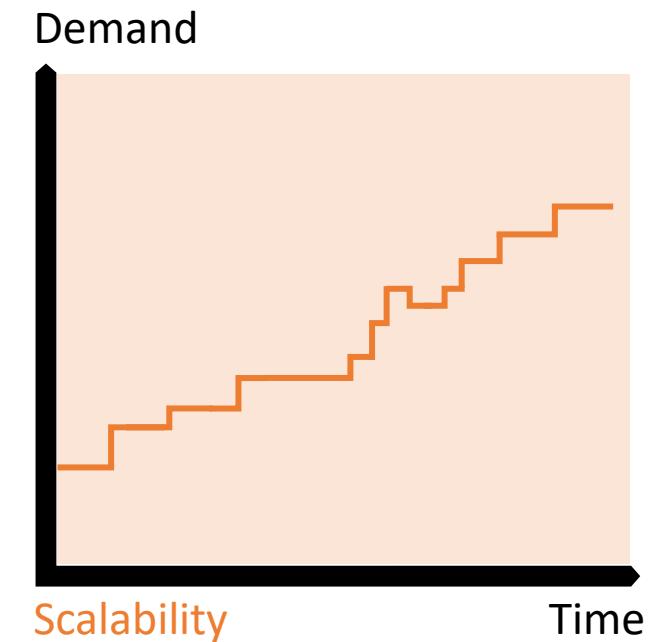
Hulk (Bruce Banner) = Scale Up



Image Credit - <https://screenrant.com/marvel-phase-5-world-war-hulk-bruce-banner/>

# What is Scalability?

- Scalability is the ability of a system to grow or shrink in a graceful manner.
- Scalability describes a design or architecture that is growth friendly.
  - Some examples
    - How efficiently a system performs when user traffic is increased multifold?
    - How well a database responds to a higher volume of queries?
    - How an operating system performs on different type of hardware?
- A scalable system performs with the same stability even if the load on it increases.



# Antman (Scott Lang) = Elasticity



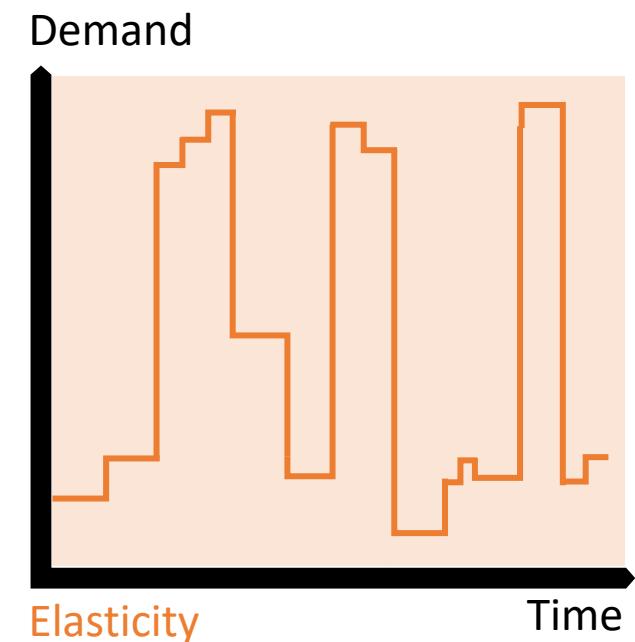
Image Credit: <https://ivleader.com/ae/2023/02/16/looking-back-at-ant-man/>

Image Credit: <https://www.themoviedb.org/collection/422834-ant-man-collection>

Image Credit: <https://insidethemagic.net/2018/06/video-new-ant-man-and-the-wasp-promo-features-a-hilarious-giant-man-prank/>

# What is Elasticity?

- Elasticity generally refers to dynamic increasing or decreasing of resources.
- An elastic system automatically adapts to match resources with demand as closely as possible, in real time.
  - Some examples
    - Adding more instances whenever a web application gets a lot of traffic.
    - Handling peak traffic like Black Friday sale
    - Scale infrastructure up for test and development activities and tear it down once test/dev work is complete
- Scalability gives you the ability to increase or decrease your resources, and elasticity lets those operations happen automatically according to configured rules.



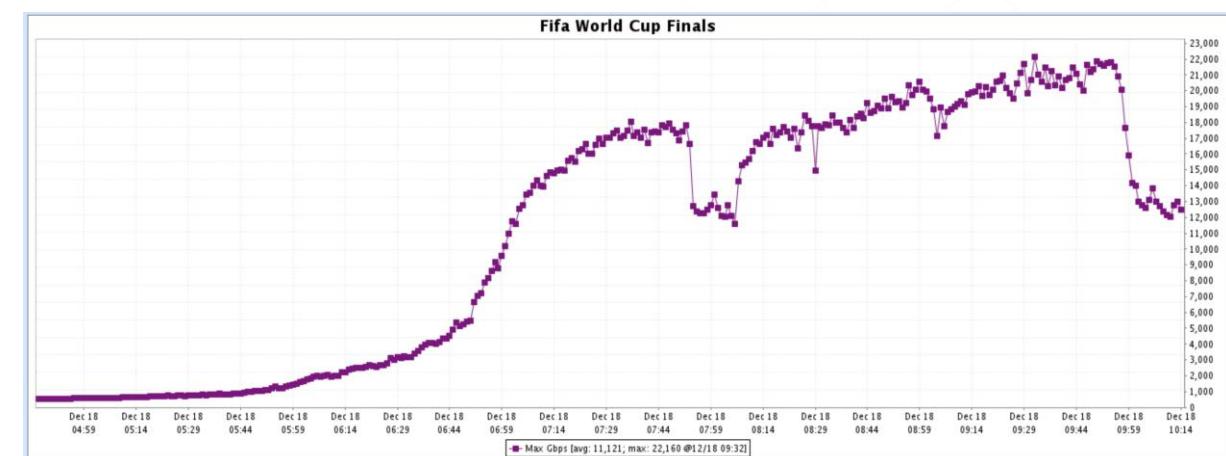
# Scalability vs. Elasticity – in Context of AWS

- Scalability – AWS expanding its global foot print

- 2006 - 2011      4 Regions
  - 2011 - 2016      7 Regions
  - 2016 - 2018      11 Regions
  - 2018 - 2023      31 Regions
  - 2008      14 Edge Locations
  - 2018      150 Edge Locations
  - 2023      400+ Edge Locations

- Elasticity – Amazon CloudFront – FIFA World Cup 2022

- Scaling up to support millions of concurrent viewers
  - Delivered more than 150 petabytes of streaming media content
  - Peaking at around 23 Terabits-per-second across
  - To 49 million unique client IP addresses

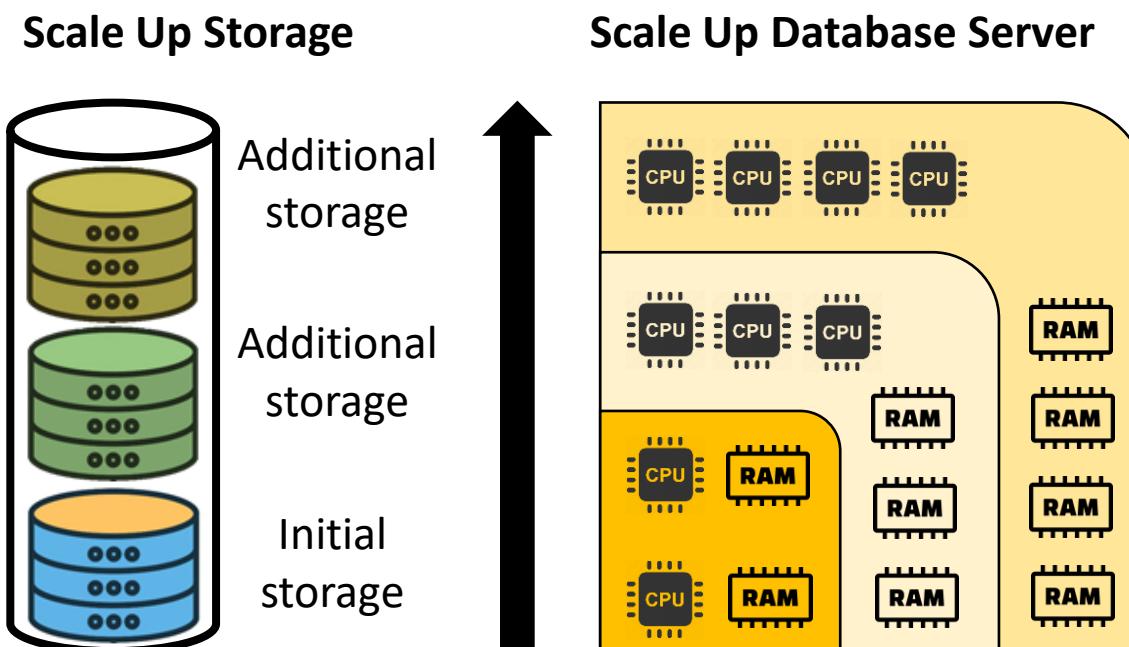


# Scalability vs. Elasticity

Scalability		Elasticity
<b>What is it?</b>	Scalability is the ability of a system to uphold the functionality when the size or volume changes.	Elasticity is the ability to dynamically manage available resources for addressing the size or volume.
<b>Use case</b>	To meet the static/predictable increase in the workload.	To meet the dynamic/sudden increase in the workload.
<b>Type</b>	Strategic operation	Tactical approach
<b>Focuses on</b>	Design/architecture	Operations
<b>Resource provisioning</b>	To exceed future demands	To meet present demand
<b>Consideration</b>	Medium- and long-term predictions	Short-term demand
<b>Execution by</b>	Typically scheduled	Typically triggered by automation

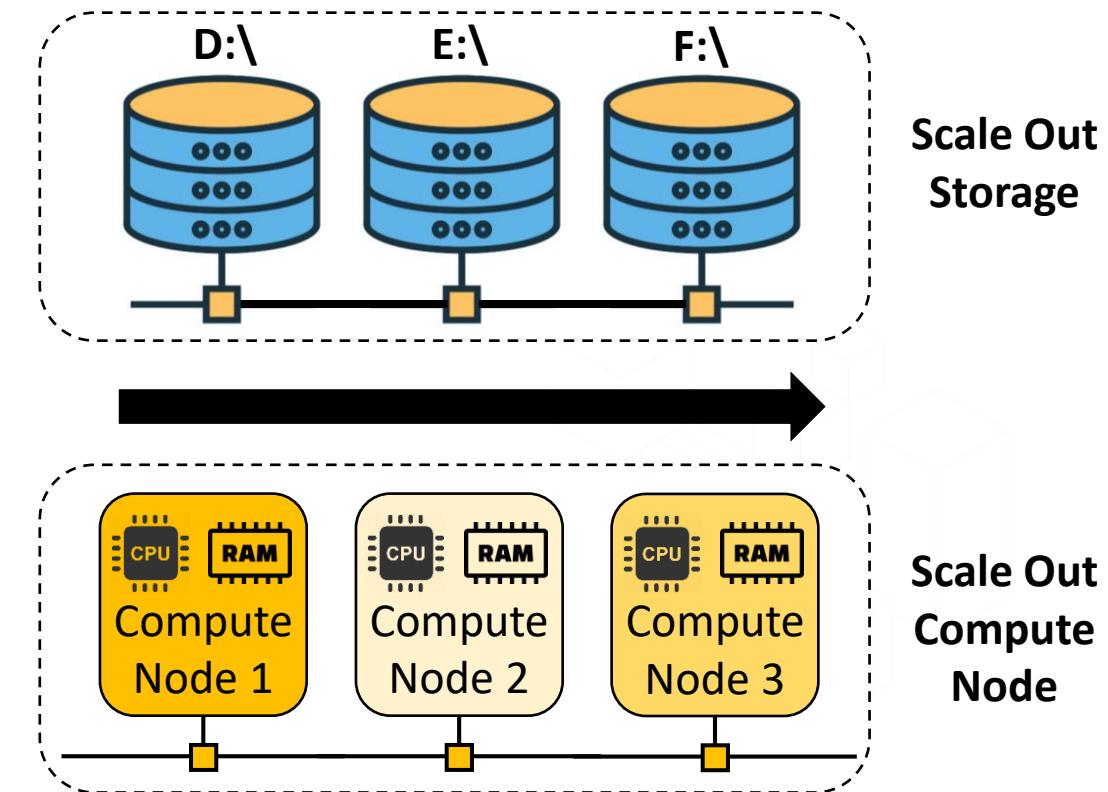
# Achieving Scalability

- Scale Up / Down (Vertical Scaling)
  - Replace the resource with a bigger size or add more capacity in the same resource
  - Suitable for – Stateful workloads



- Scale Out / In (Horizontal Scaling)

- Adding more resource to our architecture and spread the workload across those resources.
- Suitable for – Stateless workloads





**RELIABILITY**

Reliability = concept (not measured in a finite number)

- Reliability includes the ability of a system to
  - recover from infrastructure or service disruptions,
  - dynamically acquire computing resources to meet demand,
  - and mitigate disruptions such as misconfigurations or transient network issues.

# Reliability

## Reliability Design principles

- Automatically recover from failure
- Test recovery procedures
- Scale horizontally to increase aggregate workload availability
- Stop guessing capacity
- Manage change through automation

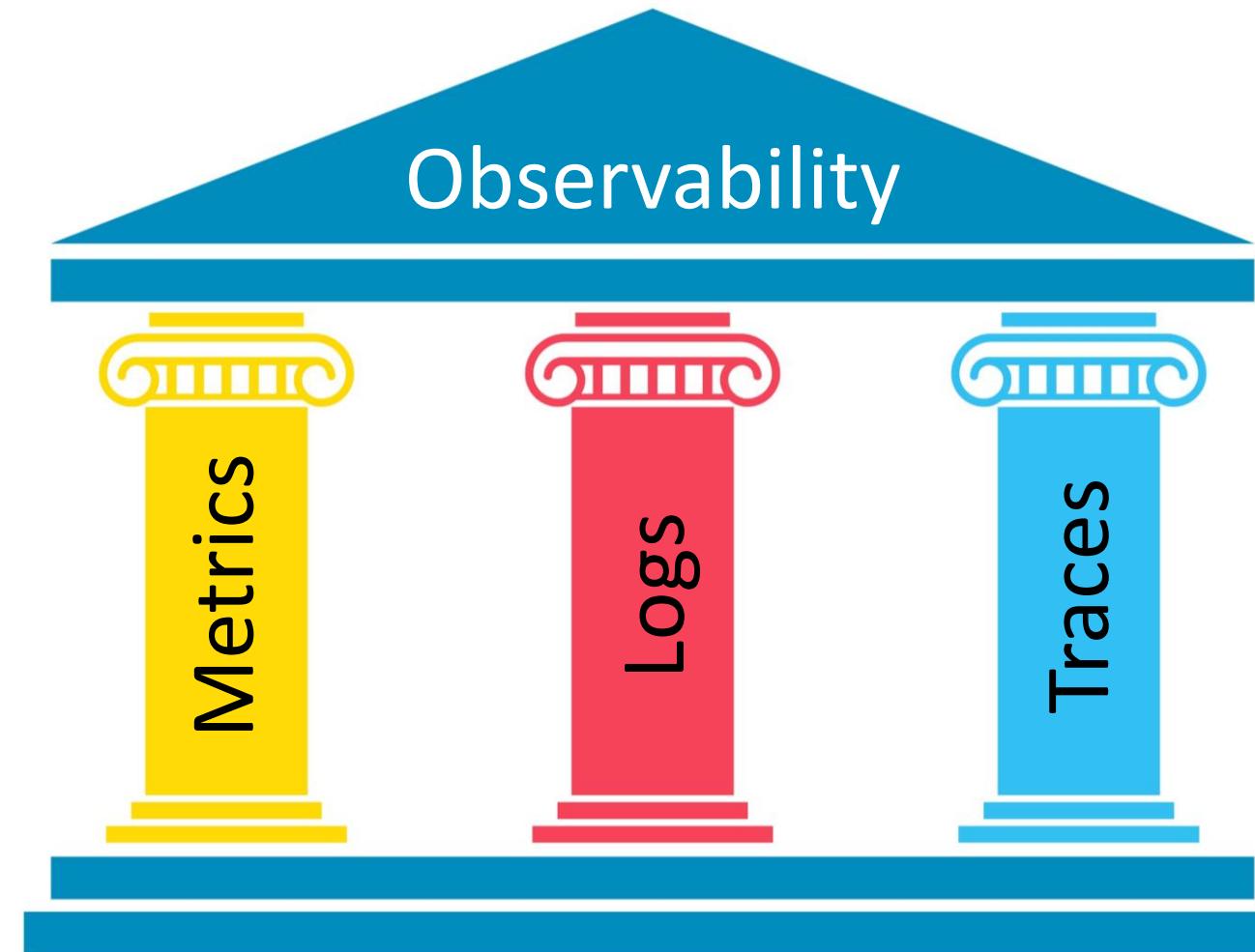
# OBSERVABILITY

# A patient under observation



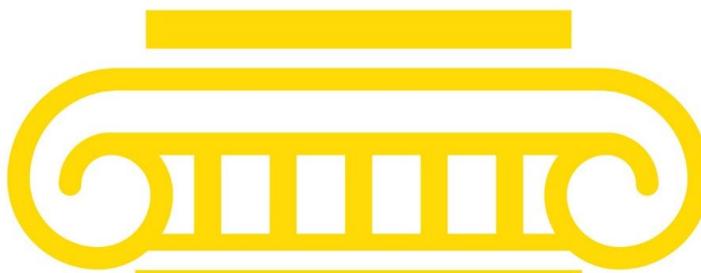
# Observability

- Observability is the extent to which you can understand the internal state or condition of a complex system based only on knowledge of its external outputs.



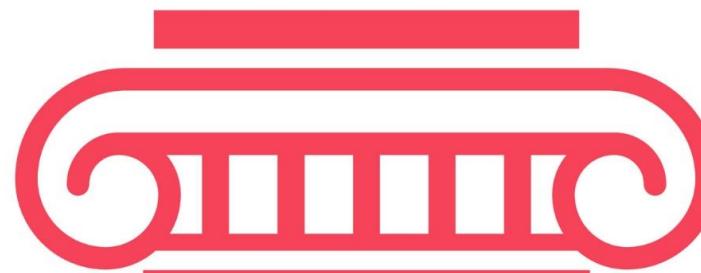
# Observability Pillars

- Observability involves gathering different types of signals and data about the components within a system, to establish the “Why?” rather than just the “What went wrong?”



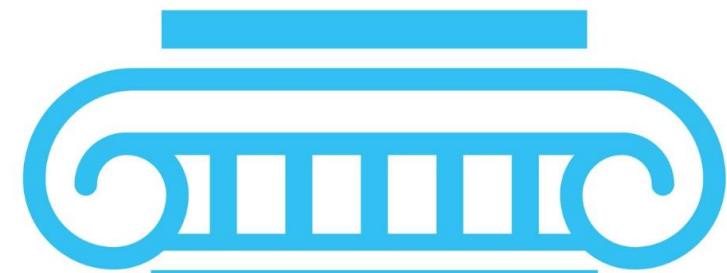
## Metrics

Time  
Series  
Data



## Logs

Sequence  
of  
Events



## Traces

End-to-end  
Request  
Flow

# How to observe?



Collect



Observe



Act



Analyze

5 Minutes  
Break

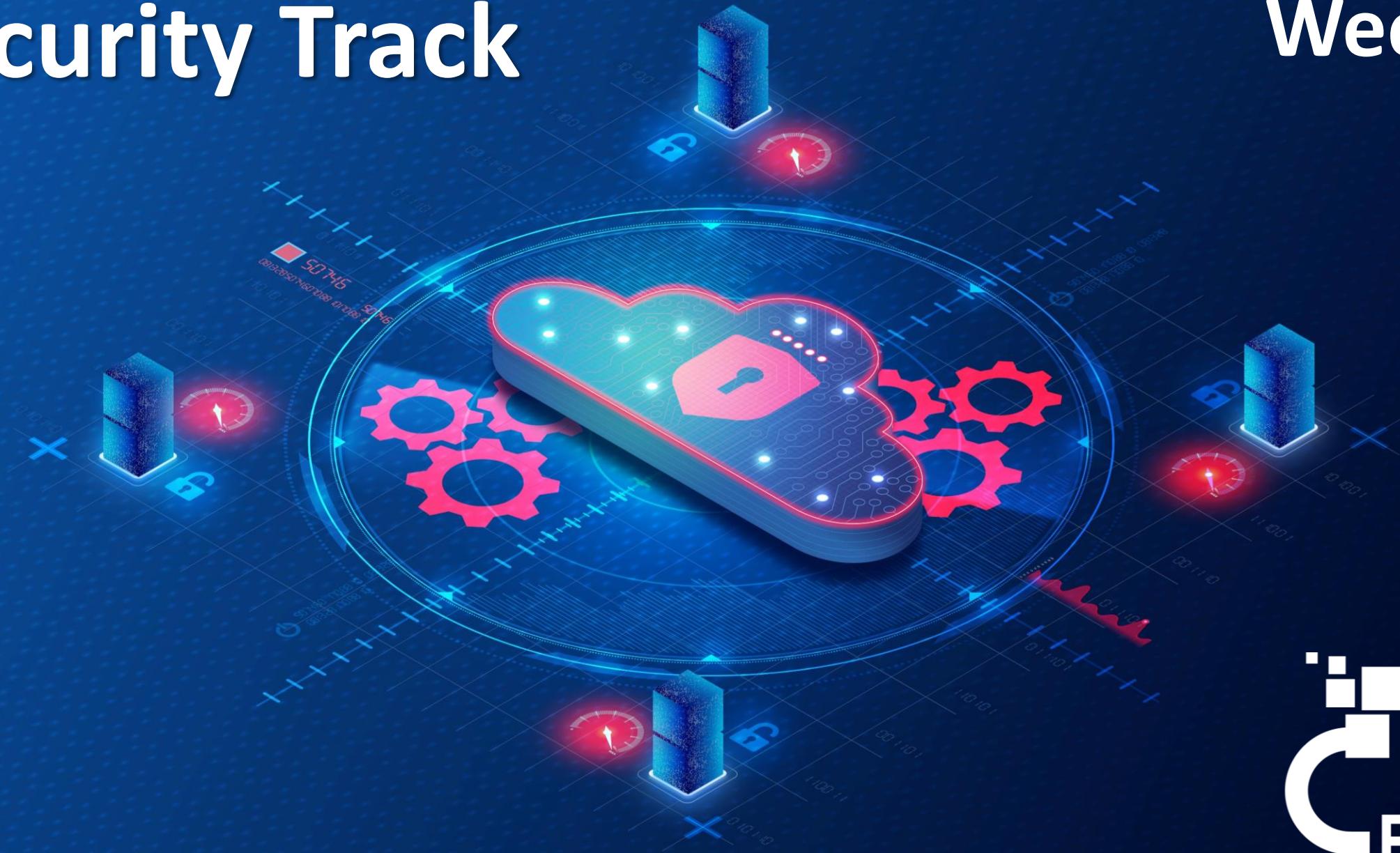


Your favourite  
AWS Service?



# Security Track

Week 4



# AWS Foundational and Layered Security Services



AWS Security Hub



AWS Organizations



AWS Control Tower



AWS Trusted Advisor



AWS Transit Gateway



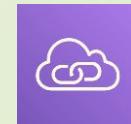
Amazon VPC



AWS IoT Device Defender



Amazon Cloud Directory



Amazon VPC PrivateLink



AWS Direct Connect



Resource Access manager



AWS Directory Service



Amazon GuardDuty



Amazon Inspector



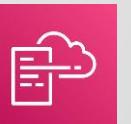
Amazon CloudWatch



AWS Step Functions



AWS OpsWorks



AWS CloudFormation

## Automate

Identify

Protect

Detect

Respond

Recover



AWS Service Catalog



AWS Config



AWS Well-Architected Tool



AWS Systems Manager



AWS Shield



IAM



AWS Secrets Manager



KMS



Amazon Cognito



AWS WAF



AWS Firewall Manager



AWS Certificate Manager



AWS CloudHSM



AWS IAM Identity Center



Amazon Macie



AWS Security Hub



Amazon Detective



Amazon CloudWatch



AWS CloudTrail



AWS Lambda



Amazon S3 Glacier



Snapshot



Archive



**Qusay Al-Maatouk**  
AWS Technical Instructor

# A day in the life of an Amazonian



**Marco Tamassia**  
AWS Technical Instructor

**Thank you for attending.  
See you next Saturday (05-Aug-2023)**



Become a Solutions Architect

For content check **Resources Link** on BeSA Home Page

