

AWS Cloud Architecture

① Cloud Terminologies

② High availability

③ High Elasticity

④ Fault tolerance

⑤ High Durability

⑥ Business continuity plan

⑦ Disaster Recovery plan

⑧ RTO

Cloud Terminologies

Solutions Architect: role that architects a technical solution using multiple systems via researching, documentation, experimentation

cloud Architect: A solutions architect that is completely focused solely on architecting technical solutions using cloud services

↳ needs to understand the following terms

① Availability → Your ability to ensure a service remains highly available

② Scalability → Your ability to grow rapidly or impeded

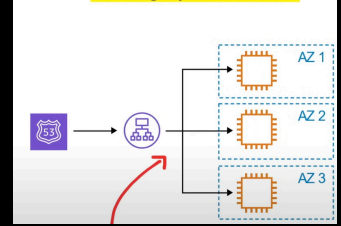
③ Elasticity → Your ability to shrink and grow to meet the demand

④ Fault tolerance → Your ability to prevent a failure

⑤ Disaster recovery → Durability

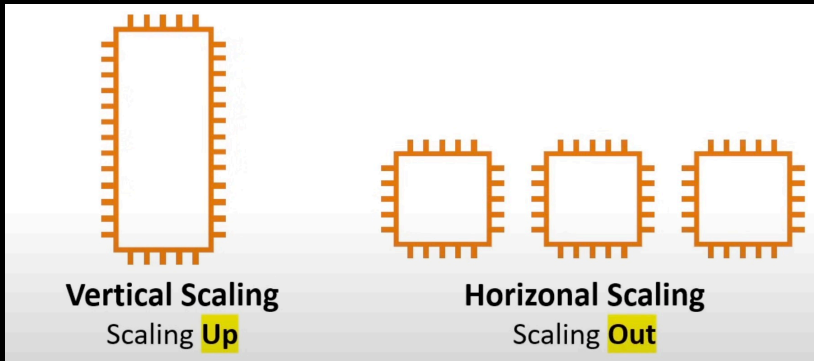
High Availability → Elastic Load Balancer

ELB: It allows you to evenly distribute traffic to multiple servers in one or more datacenters. If a data center or server becomes unavailable the load balancer will automatically route the traffic to only available data centers.



High Scalability

Ability to increase your capacity based on the increasing demand of traffic, memory & computing power



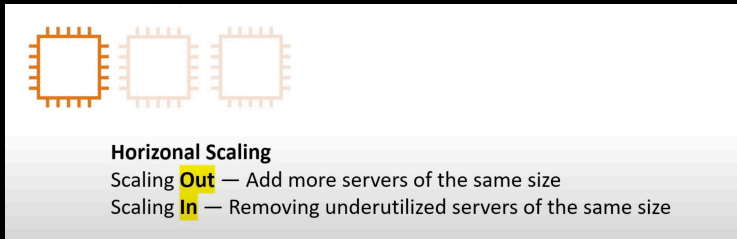
Upgrade to bigger server

Add more servers

High Elasticity → Auto Scaling Groups

Your ability to automatically increase/decrease your capacity.

* There is only horizontal scaling associated with Elasticity



ASG : is an AWS feature that will automatically add or remove servers based on scaling rules you define

High Fault tolerance → RDS Multi-AZ

RDS Multi-AZ ensures to have duplicate standby Database in case the primary DB fails.

High Durability: CloudEndure Disaster Recovery

Cloud Endure Disaster Recovery continuously replicates your machines into low cost staging area enabling fast recovery in case of IT data center failures

Business Continuity Plan

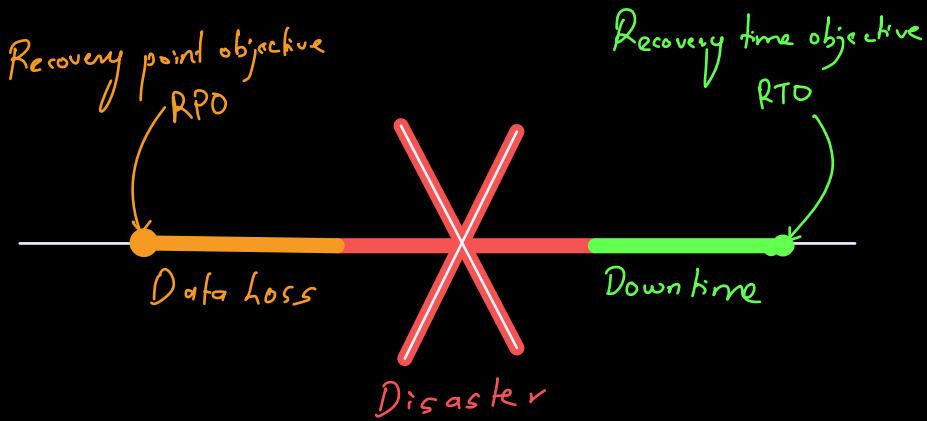
It is a document that outlines how a business will continue to operate during an unplanned disruption in services

RPO: Recovery point objective is the maximum acceptable amount of data loss after an unplanned data-loss incident expressed as an amount of time.

"How much data are you willing to lose?"

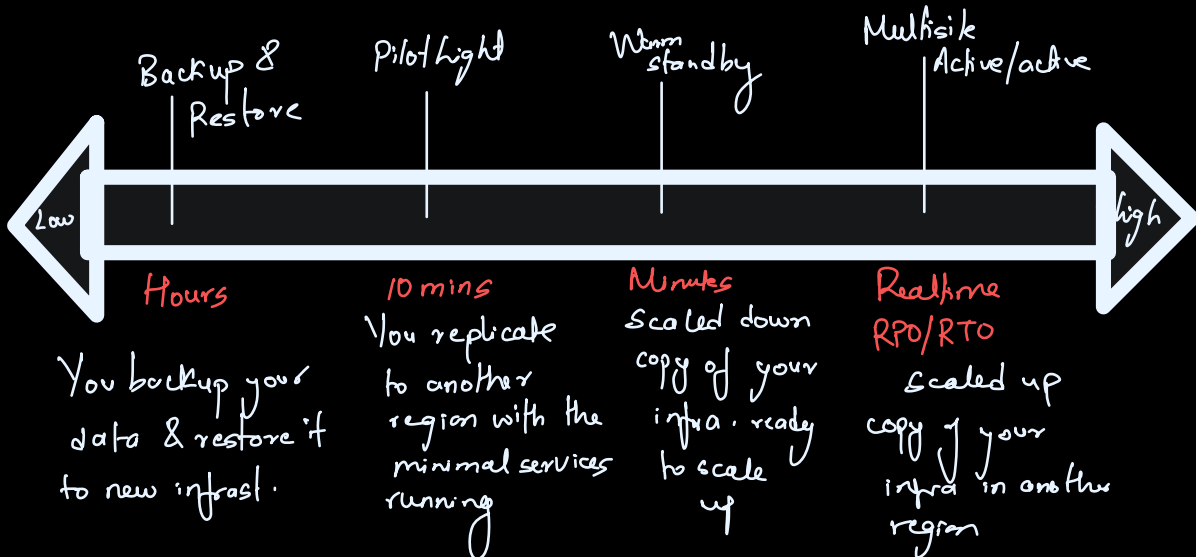
RTD: is max downtime your business can tolerate without incurring significant financial loss

"How much time are you willing to go down?"



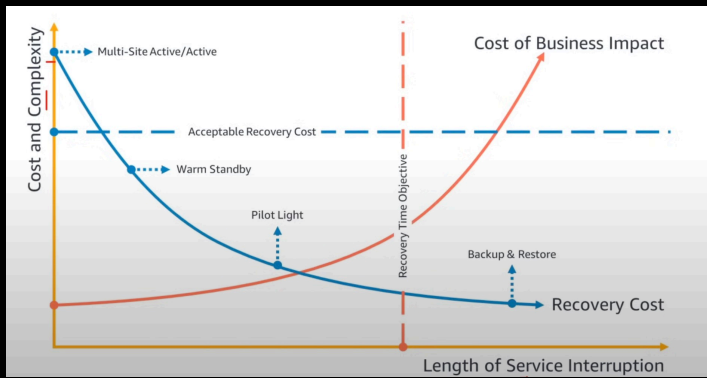
Disaster Recovery Options

It is always cost vs time where $\text{cost} \propto \frac{1}{\text{time}}$



RTO : Recovery Time Objective

Maximum delay between interruption of service and restoration of service. It determines which is an acceptable time window for an orgⁿ to have service unavailable.



RPO : Recovery Point Obj

determines what is considered an acceptable data loss between the last recovery point & interruption of service & is defined by organizations.