Terminologies

Solutinos architect- A role in a technical organization that archtiects a technical solution using multiple systems via reseacrhcing documentation, experiments.

Cloud architect-Focused solely on architectin technical solution using cloud services.

A Cloud architect needs to understand the following terms and factor them into their designed architect based on the business requirments
1) Availability
2) Scalability
3) Elasticity
4) Fault Tolerance
5) Disaster Recovery

They also need to consider Security and Cost.

High Availability- Service remain available by ensuring there is no single point of failure or ensure a certain level of performance. Eg running your workload across multiople azs. We can use **Elastic Load Blancer**. A load balancer allows you to evenly distribute traffic to multiple server in one or more datacenter, it will route the traffic to only available data center.

High Scalability- Vertical Scaling or Horizontal Scaling.

High Elasticity- Abality to automatically increase or decrease the capacity based on the demand.
Horizontal Scaling- Scaling out add more server. Scaling in-removing server
Vertical scaling is generally hard for traditional architecture so usually horizontal scaling is described with elasticity.

We can use **Auto Scaling groups (ASG)** is an AWS feture that will automatically add or remove server based on scaling rules you defined based on matrics.

Fault Tolerance- Aablity for you service to ensure there is no single point of faliure. Eg Copy of primary database which is in sync with it. For databse we can use **RDS Multi -AZ** when we run a duplicate standby datbase in another availability zone.

High Durabality-Abality to recover froma disaster to preven the loss of data. Eg we can use CloudEndure Disaster Recovery continously replicates your machines into low-cost staing are in your target aws account and preferred region enabling fast and reliable recover in case of IT data center failures.

BCP-Business continutiy plan is a document that outlines how a business will contiue during unplaned disruption in srvices.

Recovery Point Objective (RPO) the maxium acceptable amount of data loss after and unplanned data-loss incident, expressed as an amount of time

Recovery time objective (RTO) the maxium amount of downtime your business can tolerate without having a significiant financial loss.

Disaster revoery options-There are multiple options for revoery that trade cost vs time to recover.
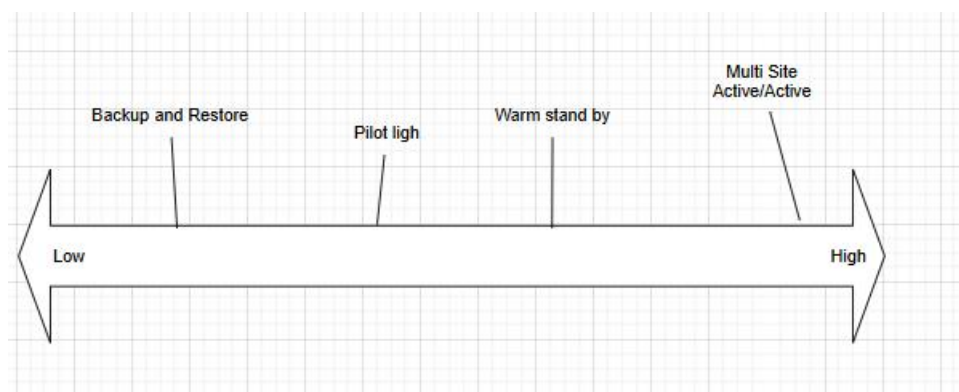
Backup& Store- You back up your data and restore it to new infa. RPO/RTO in hours.
1) Used for Low priority use cases
2) Restore data after event
3) Deploy resources after event
4) Cost $

Pilot light- Data is replicated to another region with the minimal services runnig. RPO/RTO 10 mins Data is replicated to another region with the minimal service runnig
1) Core services
2) Start and scale resources after event
3) Cost $$

Warm Standby-Scaled down copy of your infra running ready to scale up. RPO/RTO less than 10 Minutes.
1) Business critical services
2) Scal resources after event
3) Cost $$$

Multi-site- Scaled up copy for your infra. RPO/RTO real time.
1) Zero downtime
2) Near zero loss
3) Misson critical services
4) Cost $$$$ (same as infra)