

→ AWS

- Amazon web services
- It offers set of global cloud based products.
- some services provided are : computing, storage, databases, analytics, mobile developer tools, management tools, IoT, security and enterprise applications etc.

→ AWS Infrastructure

- It is built around AWS regions and availability zones.
- AWS region - a physical location in world where there are multiple availability zones.
- Availability zones - contains multiple data centres each with redundant power, networking and connectivity.
- Each Amazon region is designed to be completely isolated from each other. availability zone in a region are connected by low latency links.
- data can be stored at different geographical regions.

→ Amazon EC2

- elastic compute cloud
- IaaS
- used when we need as many, as few virtual servers, configure security and networking and managing storage
- can launch as many virtual servers as we want.
- OS - (Linux / windows / mac)
- Types
 - general purpose
 - compute optimized
 - memory optimized

→ AWS ECB

managed

- load balancer
- spreads load across multiple servers instances.
- exposes single point of access to your application
- Does regular health check of instances
- High availability across nodes.
- Separates public and private traffic
- Types
 - classic load balancer (HTTP, HTTPS, TCP)
 - Application load balancer (HTTP, HTTPS, websocket)
 - network load balancer (TCP, TCS & UDP)

→ AWS VPC

- virtual private cloud
- holds all AWS resources
- restricts traffic and IP addresses that can access instances.

→ AWS internet gateway (IGW)

- allows VPC to connect to the Internet

→ AWS subnets

- allows us to give different access rules.
- place resources in different containers where the rules should be applied.

→ AWS route tables

- contains set of rules called routes
- routes are used to direct the network traffic from subnets and gateways.

→ AWS NAT gateway

- allows instances in private subnets to connect to Internet

→ Network ACLs (NACL)

- acts like a firewall which controls traffic from and to subnets.
- deny and allow rules are used.

→ AWS security groups

- They control and allow traffic from or into of EC2 instances.
- contains rules.

→ AWS ECS

- elastic container service
- simplifies running containers across multiple capabilities.

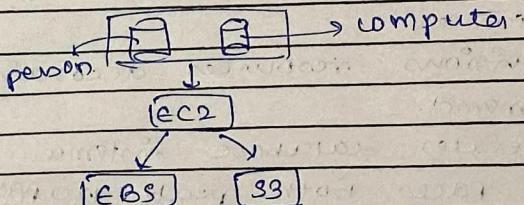
→ Terraform

- Infrastructure as code
- a tool for building, managing and versioning infrastructure safely and efficiently

→ Life cycle of web application deployment in cloud using AWS.

- Application can finally be deployed in AWS with a low cost setup and later adding features such as high performance, scalability, availability as required.

Initial deployment architecture



- a single ec2 instance is used as a starting setup to host application.
- The ec2 instance hosts both web server and database servers.

elastic block storage

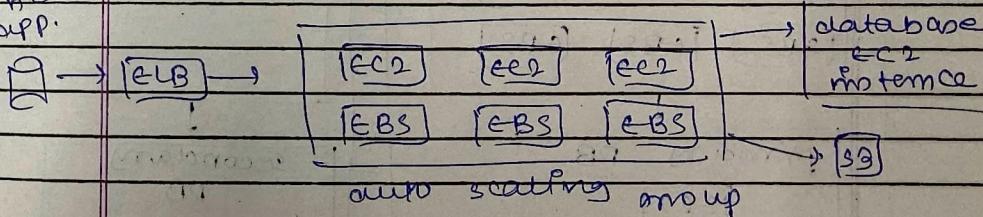
- EBS - root volume of ec2.
 - safeguards data in case ec2 fails.
 - stores application code and database files.

simple storage service

- S3 - used for storing and serving static content such as audio and video files.

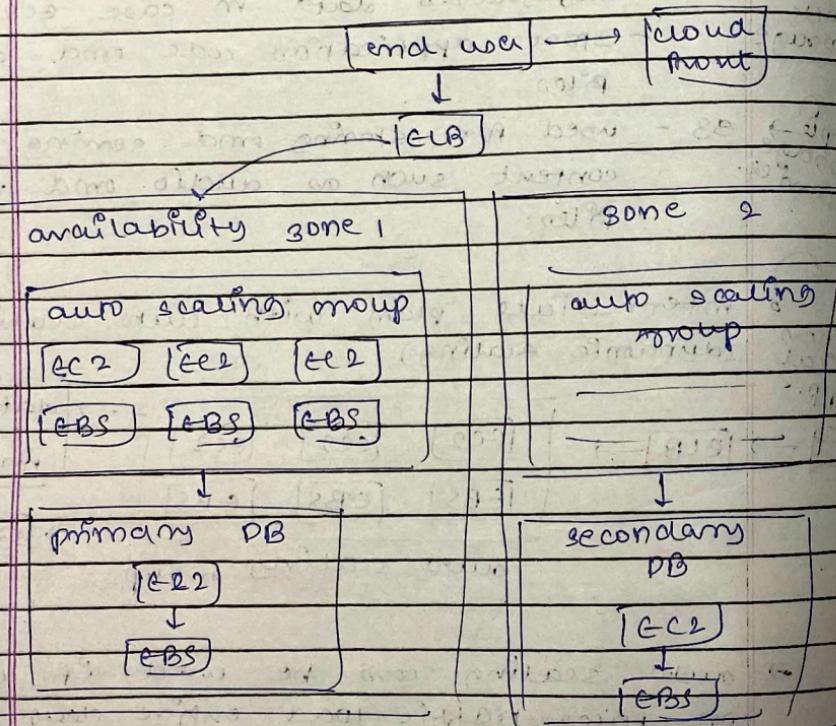
For non business general app.

Intermediate plan with auto scaling (dynamic scaling)



- auto scaling can be used if application faces traffic load entire day.
- also provides backup mechanism if a ec2 instance fails as a new instance would automatically be started.

- Here, ELB (elastic load balancer) is used to balance load evenly among different instances.
- provisions resources according to the demand.
- separates database instance.
- It faces both peak traffic and low traffic.
- Advance plan with high availability
- used for business critical applications where no downtime can be tolerated.



- adds high availability
- If a ~~instance~~ fails, application continues to be operational with slightly reduced performance. and this is done using another availability zone
- cloud front is used to cache static data close to the users.
- all availability zones must have identical configuration to ensure proper reliability.