

④

AWS

ELB & ASG Service

Elastic Load Balancing Service

Lets first understand the terms

Scalability: → Scalability means the ability to grow your system's resources when your application or website gets more traffic or more users.

Vertical Scalability (Scaling UP)

⇒ Vertical Scalability means adding more power (CPU, RAM) to your existing Server.

Ex: t2.micro to m5.large.

Horizontal Scalability (Scaling Out)

⇒ Horizontal Scalability means adding more instances (servers) to distribute the load.

⇒ You can add more EC2 instances behind (using by) a load balancer.

High Availability (HA)

⇒ High Availability (HA) means keeping your service up and running with minimal

47

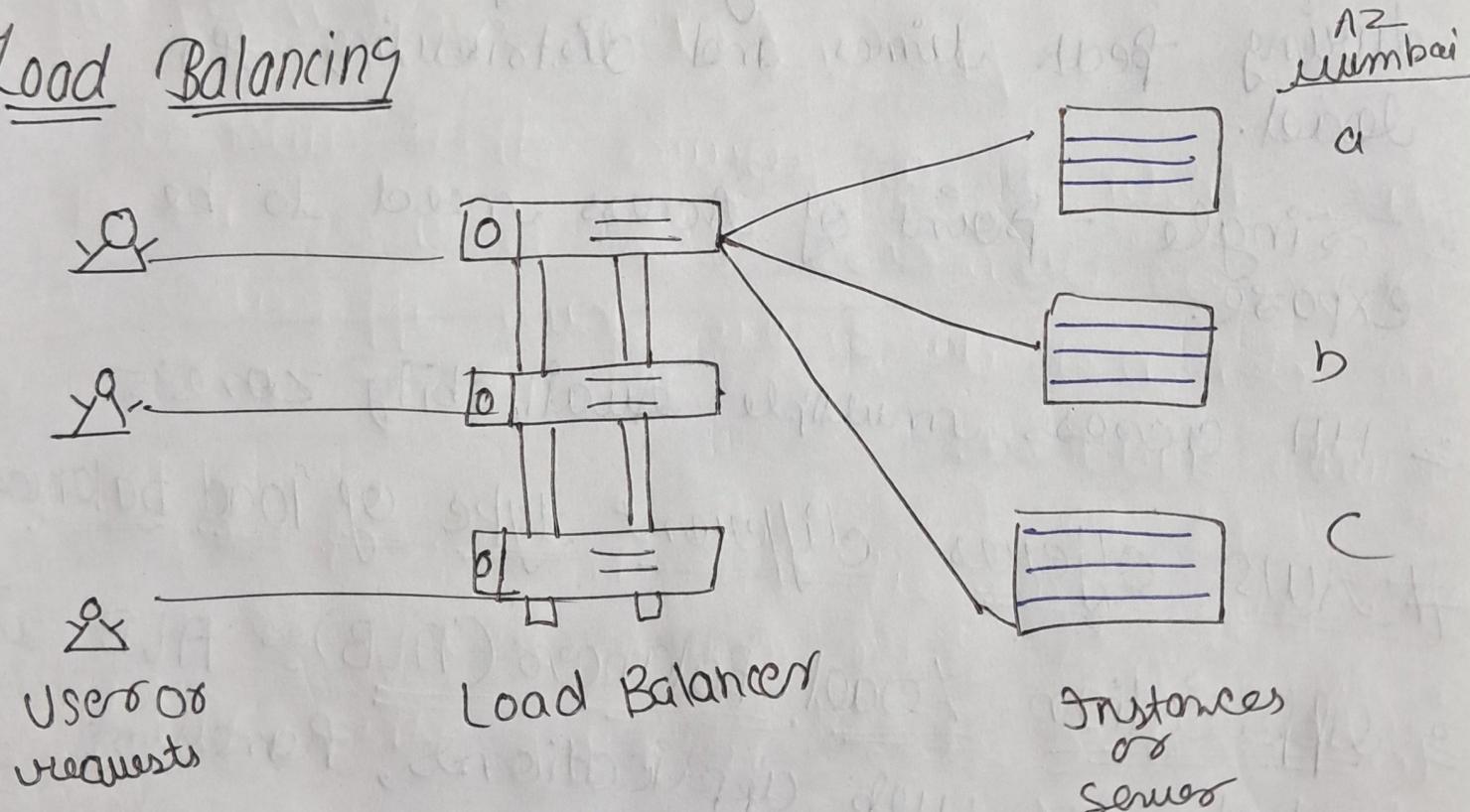
downtime, so it's always accessible to users.

Ex : running resources in multiple AZs.

⇒ Elasticity → Elasticity means the ability to automatically adjust resources as the demand changes adding more when needed and removing when it's no longer necessary.

Ex : ASG

Load Balancing



A Load Balancer in AWS automatically distributes incoming network traffic across multiple targets - like EC2 instances, containers or IP addresses - in one or more Availability Zone.

(48)

Some ELB Points

- ⇒ Distributes Traffic ⇒ It splits incoming traffic across multiple servers so no single server gets overloaded.
- ⇒ Improve Availability ⇒ If one server goes down, the load balancer automatically sends traffic to the working servers, ensuring your applications stay available.
- ⇒ Scales Resources ⇒ It helps manage high demand by adding more servers during peak times and distributing the load.
- ⇒ Single point of access need to be exposed.
- ⇒ HA across multiple availability zones.
- # AWS offers different type of load balances
- ⇒ Application Load Balancers (ALB) ⇒ ALB is perfect for web applications, handling complex HTTP and HTTPS requests (OSI Layer 7).

④ Network Load Balancer (NLB): NLB is designed for high-performance and low latency, perfect for TCP/UDP traffic (ex: gaming, financial apps) (OSI Layer 4)

⑤ Gateway Load Balancers (GWLBS): Helps deploy, scale, and manage third-party virtual applications, such as firewalls and monitoring solutions.

Steps for ELB

- ⑥ Set up EC2 instances: Create two or more EC2 instances, and launch them with user data script.
- ⑦ Configure security groups: Set up a security group allowing HTTP and SSH access.
- ⑧ Create and load balancers: Use the EC2 dashboard to create an application Load balancer.
EC2 dashboard → Load balancer → Create load balancer.

(30)

- ④ Register targets → Add EC2 instances to the target group & configure health checks.
- ⑤ Test the load balancer → Access the DNS name of the load balancer and observe load balancing in action.
- ⑥ Healthy & Unhealthy → Make sure EC2 make healthy for all.

ASG (Auto Scaling Group)

AWS ASG is a service that automatically adds or removes EC2 instances based on demand to ensure your application is always available.

It helps scale up, when more capacity is needed and scale down during low usage to save costs, keeping the right number of servers running at all times.

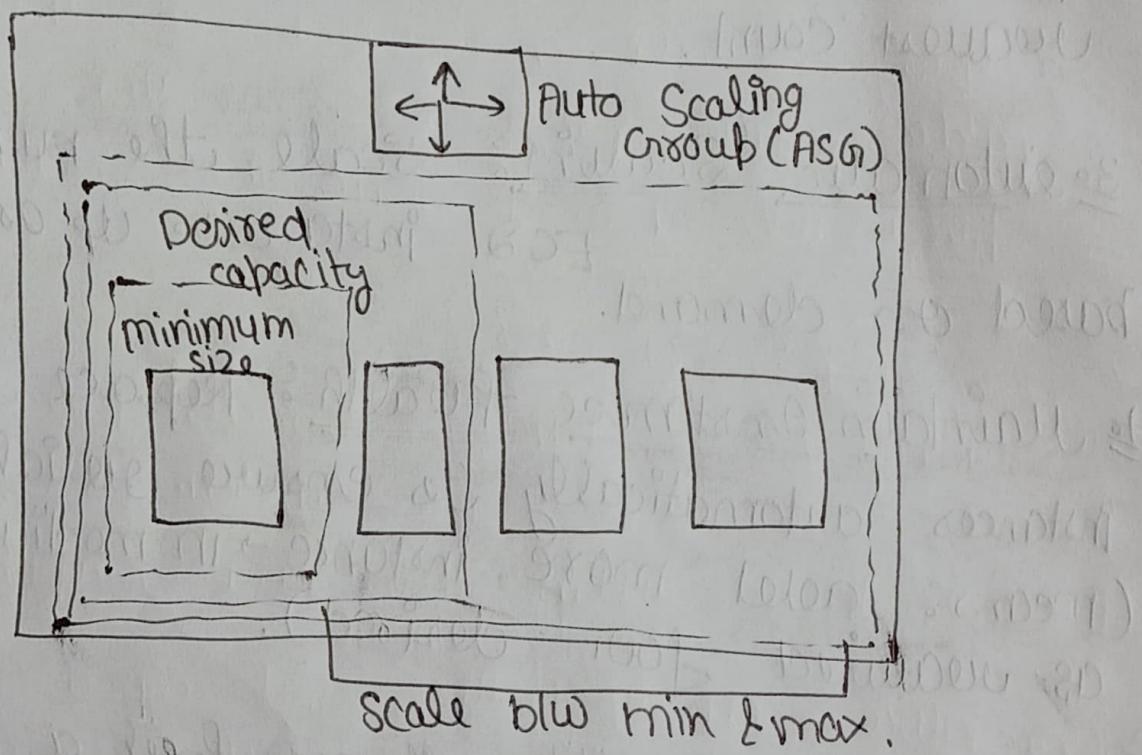
ASG Functions

1. Schedule Scaling: Pre-configure scaling activities for specific traffic peak time.
2. Use Scaling Policies: Set rules for scaling based on metrics like CPU usage or request count.
3. Automatic Scaling: Scale the number of EC2 instances up or down based on demand.
4. Maintain Instance Health: Replace unhealthy instances automatically to ensure reliability. (means: add more instance → immediately as required for desired).
5. Ensure Availability: Always keep a defined number of instances running to meet application needs.
6. Optimize Cost: Scale down during low demand to save on infrastructure costs.

1. Distribute Instances - Deploy instances across multiple availability zones for high availability.

2. Integrate with ELB (Elastic Load Balancing)

Attach instances to an ELB to automatically balance traffic.



3. Steps to Create ASG

1. Create Launch Template - define instance settings (AMI, type, key etc.).

2. Create ASG → Start creating the Auto Scaling group.

3. Select VPC & Subnet → choose the network for your instances.

4. Attach load balancers (optional) - distribute traffic across instances.
(Recommended)
5. Set Scaling Policies :- Define when to scale in/out.
6. Configure health checks :- EC2 or ELB-based monitoring.
7. Enable Notifications (optional) :- Get alert for ASG
8. Review & Launch :- Confirm settings & create the ASG.

Instances | EC2 | ap-south-1 ChatGPT elb aws - Google Search

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#Instances:

aws | Instances | Search [Alt+S]

EC2 Instances

EC2 Dashboard EC2 Global View Events

Instances Instances Instance Types Launch Templates Spot Requests Savings Plans Reserved Instances Dedicated Hosts Capacity Reservations

Images AMIs AMI Catalog

Elastic Block Store Volumes Snapshots Lifecycle Manager

CloudShell Feedback

Last updated less than a minute ago Connect Instance state Actions Launch instances

Find Instance by attribute or tag (case-sensitive) All states

No instances You do not have any instances in this region Launch instances

Select an instance

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AIRTELPP +1.07% 12:22 23-07-2025

Search

Launch an instance | EC2 | ap-south-1 ChatGPT elb aws - Google Search

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#LaunchInstances:

aws Search [Alt+S] EC2 Instances Launch an instance

Launch an instance Info

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags Info

Name
instance

Add additional tags

Application and OS Images (Amazon Machine Image) Info

An AMI contains the operating system, application server, and applications for your instance. If you don't see a suitable AMI below, use the search field or choose **Browse more AMIs**.

Search our full catalog including 1000s of application and OS images

Quick Start

Amazon Linux macOS Ubuntu Windows Red Hat SUSE Linux Debian

AWS Mac ubuntu Microsoft Red Hat SUSE Linux debian

Browse more AMIs
Including AMIs from AWS, Marketplace and the Community

Summary

Number of instances Info
2

When launching more than 1 instance, consider **EC2 Auto Scaling**

Firewall (security group)
New security group

Storage (volumes)
1 volume(s) - 8 GiB

Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the internet.

Cancel Launch instance Preview code

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AIRTELPP +1.07% 12:23 23-07-2025

Load balancer details | EC2 | EC2 Instance Connect | ap- | EC2 Instance Connect | ap- | Add and run script | elb aws - Google Search | 13.201.129.135 | + | - | X

← → C ap-south-1.console.aws.amazon.com/ec2-instance-connect/ssh/home?region=ap-south-1&connType=standard&instanceId=i-096985872f60e6507&osUser=ec2-user&sshPort=22&addr... ☆ | ☰ | ↴ | ⌂ | Y | ...

aws | ☰ | Q Search [Alt+S] | ☰ | ☰ | ☰ | ☰ | Asia Pacific (Mumbai) | Yash Garg | ☰

GNU nano 8.3 server.js Modified

```
const http = require('http');
const os = require('os');

const hostname = os.hostname(); // Get the machine's hostname

const server = http.createServer((req, res) => {
  if (req.url === '/health') {
    res.writeHead(200, { 'Content-Type': 'text/plain' });
    res.end('OK');
    return;
  }

  res.writeHead(200, { 'Content-Type': 'text/html' });
  res.end(`<h1>Hello from ${hostname}</h1>`);
});

// Listen on all network interfaces (important!)
server.listen(80, '0.0.0.0', () => {
  console.log(`Server running at http://${hostname}:80/`);
});
```

^G Help ^O Write Out ^F Where Is ^K Cut ^T Execute ^C Location M-U Undo M-A Set Mark M-] To Bracket M-B Previous
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^/ Go To Line M-E Redo M-6 Copy ^B Where Was M-F Next

i-096985872f60e6507 (instance)

PublicIPs: 13.201.129.135 PrivateIPs: 172.31.10.103

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

32°C Cloudy Search ☰ PRE 🔍 📁 🚙 🎯 🎮 🎵 🎥 🎨 🎪 🎧 🎧 ENG IN 13:19 23-07-2025

Instance details | EC2 | ap-south-1 | EC2 Instance Connect | ap-south-1 | Add and run script | elb aws - Google Search | 13.201.129.135 | +

← → C ap-south-1.console.aws.amazon.com/ec2-instance-connect/ssh/home?addressFamily=ipv4&connType=standard&instanceId=i-096985872f60e6507&osUser=ec2-user®ion=ap-south-1 | Search [Alt+S] | AWS | ☰ | ⓘ | ⓘ | ⓘ | Asia Pacific (Mumbai) | Yash Garg | ☰

```
/m/ /  
[ec2-user@ip-172-31-10-103 ~]$ ls  
[ec2-user@ip-172-31-10-103 ~]$ lsblk  
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINTS  
xvda 202:0 0 8G 0 disk  
└─xvda1 202:1 0 8G 0 part /  
└─xvda127 259:0 0 1M 0 part /  
└─xvda128 259:1 0 10M 0 part /boot/efi  
[ec2-user@ip-172-31-10-103 ~]$ sudo dnf install -y nodejs npm
```

node -v
Amazon Linux 2023 Kernel Livepatch repository
Dependencies resolved.

Package	Architecture	Version	Repository	Size
Installing:				
nodejs	x86_64	1:18.20.8-1.amzn2023.0.1	amazonlinux	13 M
nodejs-npm	x86_64	1:10.8.2-1.18.20.8.1.amzn2023.0.1	amazonlinux	1.9 M
Installing dependencies:				
libbrotli	x86_64	1.0.9-4.amzn2023.0.2	amazonlinux	315 k
nodejs-libs	x86_64	1:18.20.8-1.amzn2023.0.1	amazonlinux	14 M
Installing weak dependencies:				
nodejs-docs	noarch	1:18.20.8-1.amzn2023.0.1	amazonlinux	7.8 M
nodejs-full-i18n	x86_64	1:18.20.8-1.amzn2023.0.1	amazonlinux	8.4 M

i-096985872f60e6507 (instance)
PublicIPs: 13.201.129.135 PrivateIPs: 172.31.10.103

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

INFY +1.09% 12:37 23-07-2025

Load balancers | EC2 | ap-south-1 | EC2 Instance Connect | ap-south-1 | EC2 Instance Connect | ap-south-1 | Add and run script | elb aws - Google Search | 13.201.129.135

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#LoadBalancers:

aws | Search [Alt+S]

EC2 > Load balancers

AMI Catalog

Elastic Block Store

- Volumes
- Snapshots
- Lifecycle Manager

Network & Security

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

Load Balancing

- Load Balancers**
- Target Groups
- Trust Stores

Auto Scaling

- Auto Scaling Groups

Settings

Load balancers

Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

Filter load balancers

Name State Type Scheme IP address type VPC ID Availability Zone

No load balancers

You don't have any load balancers in ap-south-1

Create load balancer

Actions

Create load balancer

CloudShell Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

6 32°C Cloudy

Search

12:57 23-07-2025 ENG IN

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#CreateSecurityGroup

aws | [Alt+S]

EC2 > Security Groups > Create security group

Outbound rules

Type	Info	Protocol	Info	Port range	Info	Destination	Info	Description - optional	Info
HTTP	▼	TCP		80		Custom	▼	0.0.0.0/0	X

Add rule

Rules with destination of 0.0.0.0/0 or ::/0 allow your instances to send traffic to any IPv4 or IPv6 address. We recommend setting security group rules to be more restrictive and to only allow traffic to specific known IP addresses.

Tags - optional

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

No tags associated with the resource.

Add new tag

You can add up to 50 more tags

Cancel Create security group

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

32°C Cloudy ENG IN 13:07 23-07-2025

Instances | EC2 | ap-south-1 | EC2 Instance Connect | ap-south-1 | Add and run script | elb aws - Google Search | 13.201.129.135 | +

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#Instances:

Search [Alt+S]

EC2 Instances

AMI Catalog

Elastic Block Store

- Volumes
- Snapshots
- Lifecycle Manager

Network & Security

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

Load Balancing

- Load Balancers
- Target Groups
- Trust Stores

Auto Scaling

- Auto Scaling Groups
- Settings

Last updated 3 minutes ago

Connect

Instance state ▾

Actions ▾

Launch instances ▾

Find Instance by attribute or tag (case-sensitive)

All states ▾

Name	Instance ID	Instance state	Instance type	Status check	Alarm s
<input checked="" type="checkbox"/> instance	i-096985872f60e6507	Running	t2.micro	2/2 checks passed	View all
<input type="checkbox"/> instance	i-0a6a16abec01ba5a	Running	t2.micro		

Create image

Create template from instance

Launch more like this

i-096985872f60e6507 (instance)

Details Status and alarms Monitoring Security Networking Storage Tags

Instance summary

Instance ID: i-096985872f60e6507

IPv6 address: -

Public IPv4 address: 13.201.129.135 | [open address](#)

Private IPv4 addresses: 172.31.10.103

Instance state: Running

Public DNS: ec2-13-201-129-135.ap-south-1.compute.amazonaws.com | [open address](#)

CloudShell Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

32°C Cloudy

Search

12:42 23-07-2025 ENG IN

Create application load balancer Step 1 Create target group EC2 Instance Connect EC2 Instance Connect Add and run script elb aws - Google Search 13.201.129.135

ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#CreateTargetGroup:

aws | Search [Alt+S] | ☰ | 🔍 | ⓘ | ⓘ | ⓘ | Asia Pacific (Mumbai) | Yash Garg | ☰ | ☰

EC2 > Target groups > Create target group

Step 1 Specify group details

Step 2 Register targets

Specify group details

Your load balancer routes requests to the targets in a target group and performs health checks on the targets.

Basic configuration

Settings in this section can't be changed after the target group is created.

Choose a target type

Instances

- Supports load balancing to instances within a specific VPC.
- Facilitates the use of [Amazon EC2 Auto Scaling](#) to manage and scale your EC2 capacity.

IP addresses

- Supports load balancing to VPC and on-premises resources.
- Facilitates routing to multiple IP addresses and network interfaces on the same instance.
- Offers flexibility with microservice based architectures, simplifying inter-application communication.
- Supports IPv6 targets, enabling end-to-end IPv6 communication, and IPv4-to-IPv6 NAT.

Lambda function

- Facilitates routing to a single Lambda function.
- Accessible to Application Load Balancers only.

Application Load Balancer

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

6 32°C Cloudy Search ☰ PRE 🔍 📁 🌐 🎯 🎮 🎵 🎥 🎪 🎧 ENG IN 12:57 23-07-2025

 EC2 > Target groups > instance-tg

instance-tg

Actions ▾

Details					
arn:aws:elasticloadbalancing:ap-south-1:654049343667:targetgroup/instance-tg/415ca2f1b65ad1d6					
Target type	Protocol : Port	Protocol version	VPC		
Instance	HTTP: 80	HTTP1	vpc-Of4849a1ce9840022		
IP address type	Load balancer				
IPv4	None associated				
2 Total targets	0 Healthy	0 Unhealthy	2 Unused	0 Initial	0 Draining
	0 Anomalous				

► Distribution of targets by Availability Zone (AZ)

Select values in this table to see corresponding filters applied to the Registered targets table below.

Targets

Monitoring

Health checks

Attributes

Tags

Registered targets (2) Info

(i) Anomaly mitigation: Not applicable

Deregister

Register targets

Target groups route requests to individual registered targets using the protocol and port number specified. Health checks are performed on all registered targets according to the target



Services

Search

[Option+S]



Stock

EC2 Dashboard



EC2 Global View

Events

▼ Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations New

▼ Images

AMIs

AMI Catalog

▼ Elastic Block Store



Listeners and rules

Network mapping

Resource map - new

Security

Monitor

Resource map Info

View, explore, and troubleshoot your load balancer's architecture.

Overview

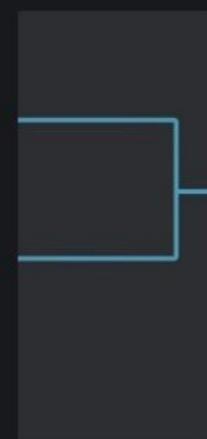
Unhealthy target map



Show resource details

mywebserver-lb

Last fetched seconds ago

Target groups (1) Info

Instance

2 targets | 0 unhealthy

mywebserver-lb-tg



2



0



0



0



0

Targets (2)

i-09185eb0a8bd41ad8

Healthy

i-0ba05e7362037134e

Healthy