

Scalable Computing in AWS



Ryan Lewis

CLOUD ENGINEER

@ryanmurakami ryanlewis.dev



Meet User Demands with Scaling

Overview

Scalability !== Elasticity

Launching a launch configuration

Balancing requests with an ELB

Scale in your sleep

When to scale?

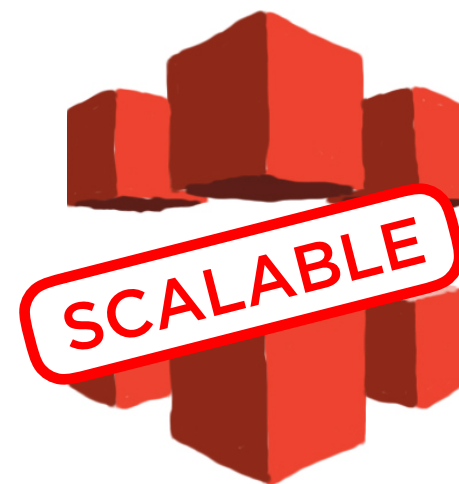
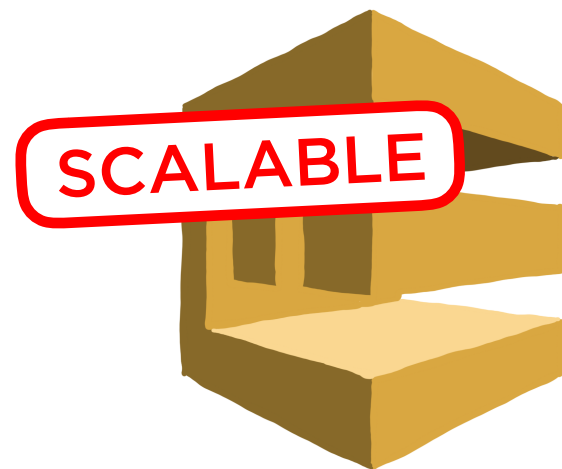
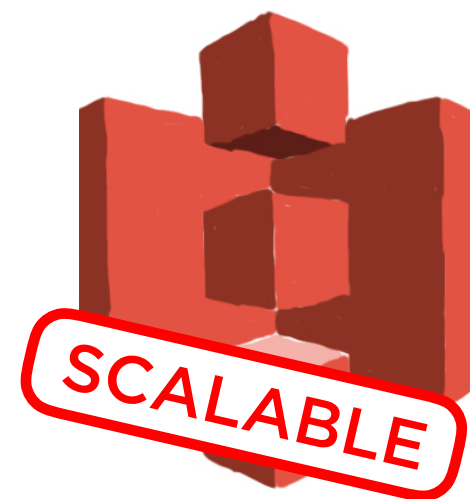
Limits to the scaling magic

Understanding Scalability and Elasticity with AWS

Scalability \neq Elasticity

Scalability

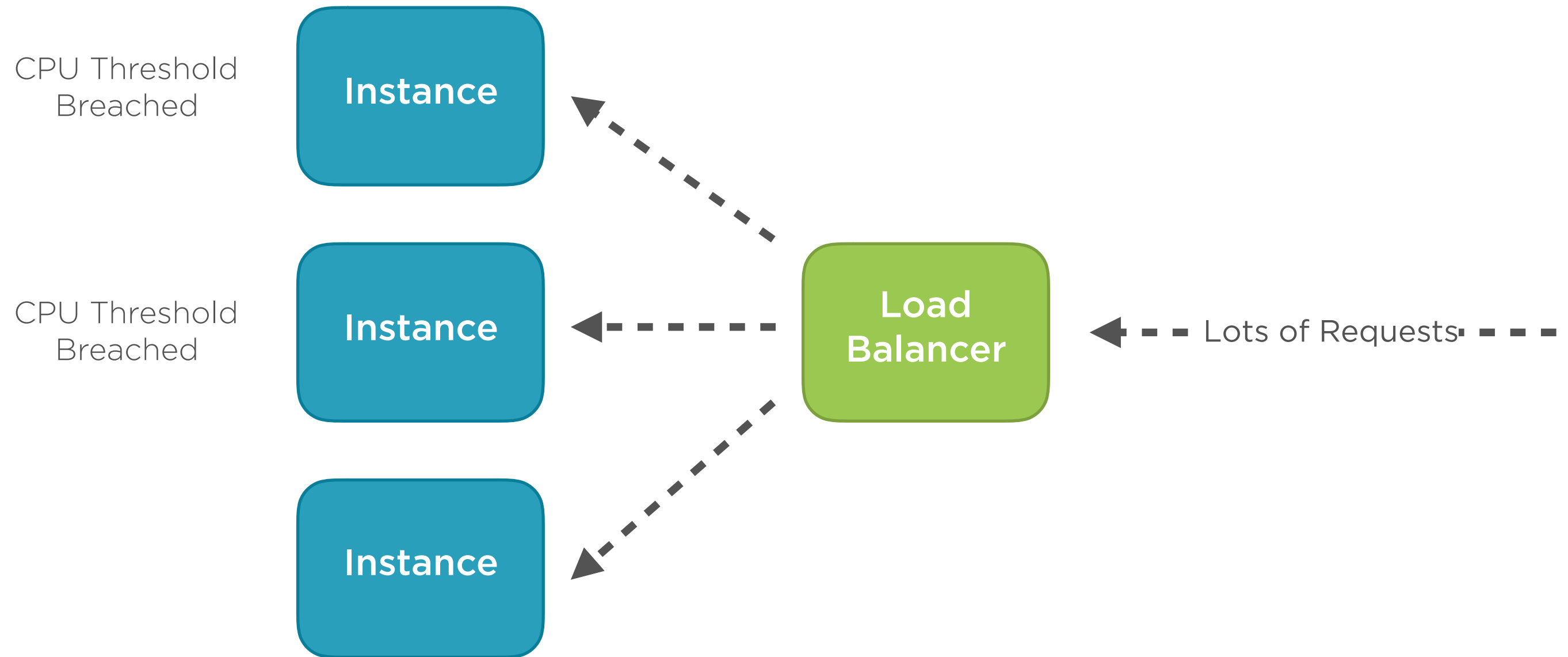
Increasing or decreasing the size or quantity of a resource in AWS.



Elasticity

Scaling in response to preset rules, often triggered by CloudWatch alarms.

Elasticity Example



Scalability

+

=



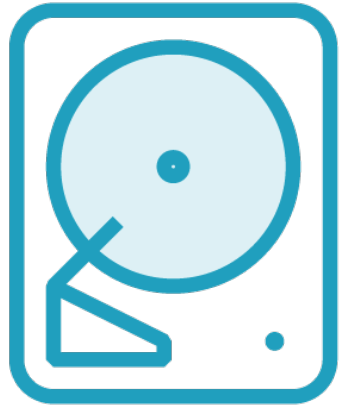
Elasticity

Creating a Launch Configuration

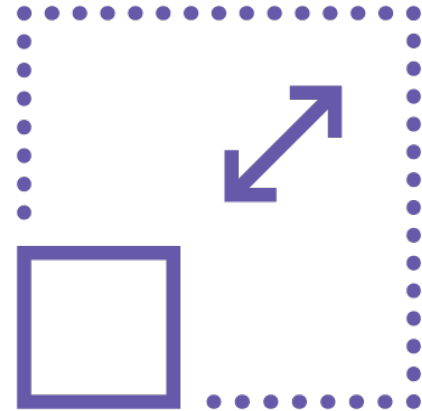
Launch Configuration

Blueprint for creating an EC2 instance.

Launch Configuration Attributes



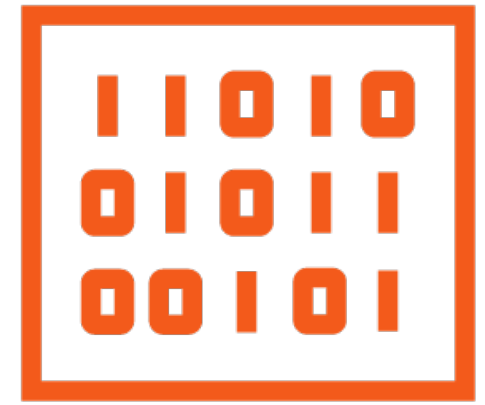
AMI



Instance type



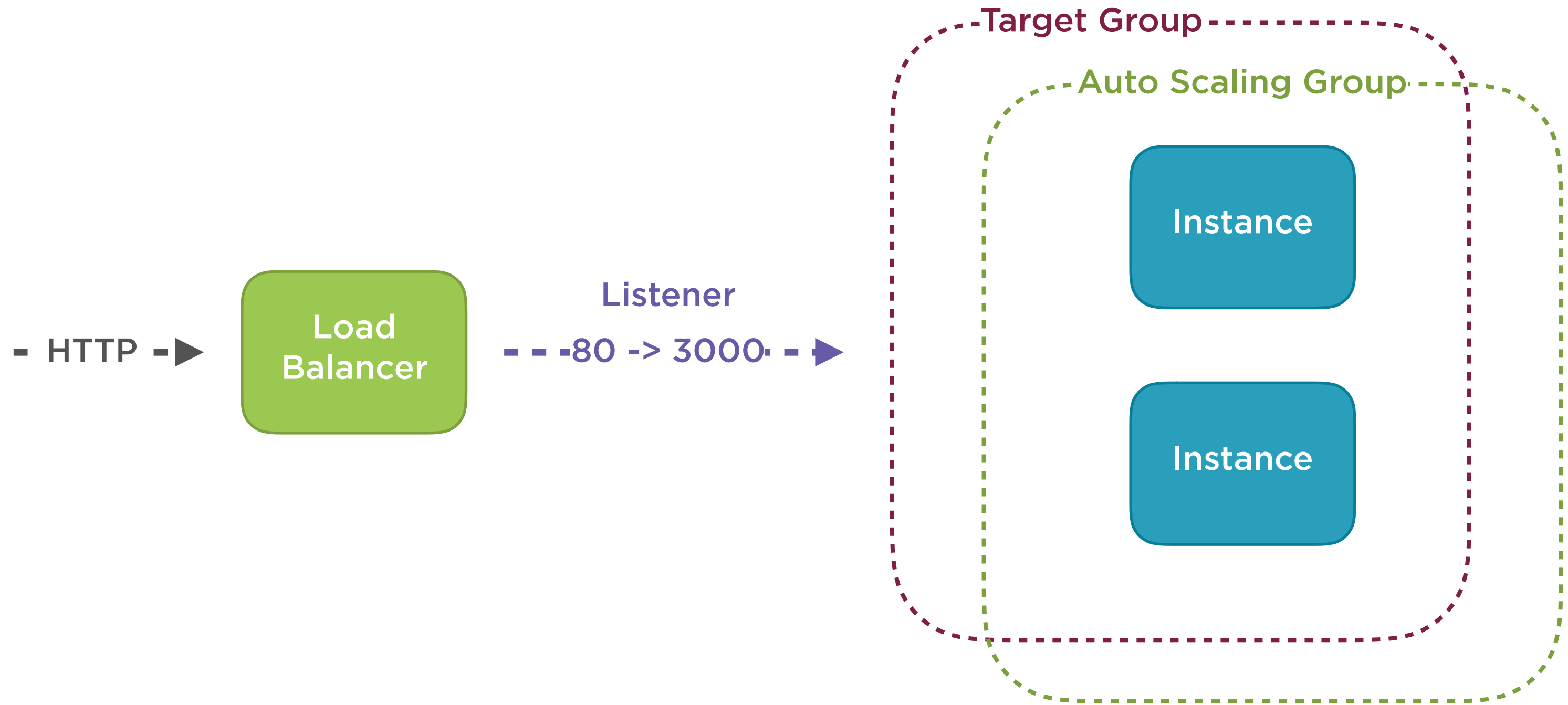
Security groups



User data

Auto Scaling groups depend
on launch configurations

Creating an Elastic Load Balancer



Creating an Auto Scaling Group

Configuring Auto Scaling Policies



Simple Scaling Policy

Monitor attributes and perform action

One action per alarm

Scaling up and down requires two policies



Step Scaling Policy

Define multiple actions per alarm

Continuously perform actions



Target Tracking Scaling Policy

Define metric target

Auto Scaling up and down to achieve target

AWS recommended scaling policy

Limits with Auto Scaling and ELB

EC2 Limits Page

EC2 Management Console

+

←

→

↺

🏠

https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Limits:

...

🔒

☆

☰

📄

👤

☰

aws

Services ▾

Resource Groups ▾

📦 Lambda

📦 EC2

📦 S3

📦

🔔

Ryan H Lewis ▾

N. Virginia ▾

Support ▾

🔵

New EC2 Experience

Tell us what you think

EC2 Dashboard New

Events

Tags

Reports

Limits

▼ INSTANCES

Instances

Instance Types

Launch Templates New

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Scheduled Instances

Capacity Reservations

▼ IMAGES

AMIs

☐

General Purpose (SSD) volume storage (TiB)

EBS

300

☐

Provisioned IOPS (SSD) volume storage (TiB)

EBS

300

☐

Provisioned IOPS

EBS

300000

☐

Max Throughput Optimized HDD (ST1) Storage...

EBS

300

☐

Max Cold HDD (SC1) Storage in TiB

EBS

300

☐

Magnetic volume storage (TiB)

EBS

300

☐

Classic Load Balancers

Load balancing

20

☐

Target groups

Load balancing

3000

☐

Network Load Balancers

Load balancing

50

☐

Application Load Balancers

Load balancing

50

☐

EC2-Classic Elastic IPs

Networking

5

☐

Subnets per VPC

Networking

200

☐

Network interfaces

Networking

5000

☐

VPCs

Networking

5

🗨 Feedback

🌐 English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

Auto Scaling Group Limit

Soft limit on number of groups and launch configurations

Elastic Load Balancing Limit

Soft limit on number of application/network load balancers

Elastic Load Balancing Limit

One SSL Certificate per load balancer

Elastic Load Balancing Limit

One target group per load balancer

Conclusion

Summary

Scalability or elasticity? Why not both?!

A configuration to launch from

Who balances the load balancers?

Automobile skilling grape

ASG policy dictatorship

Mind the resource limits

Up Next

Storage in AWS