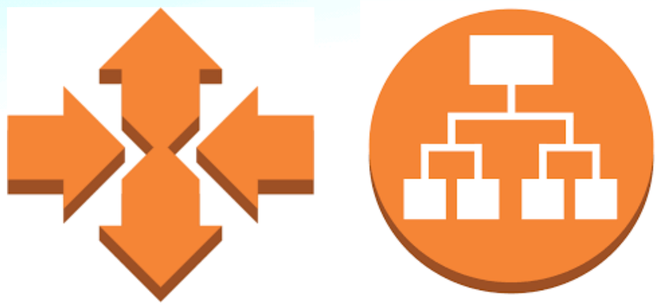
Auto Scaling Group and Load Balancer

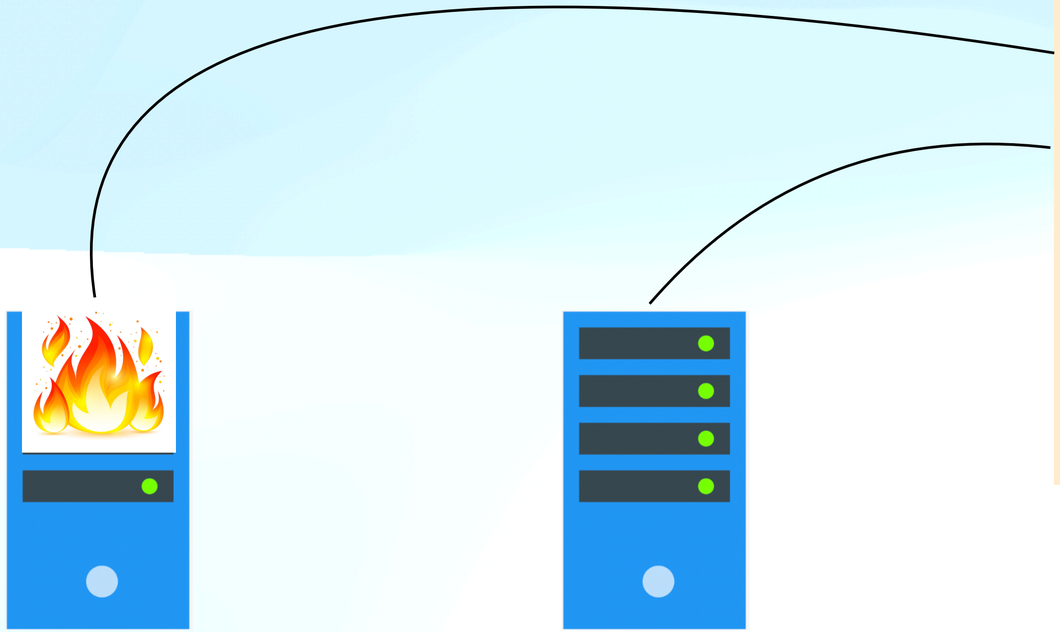


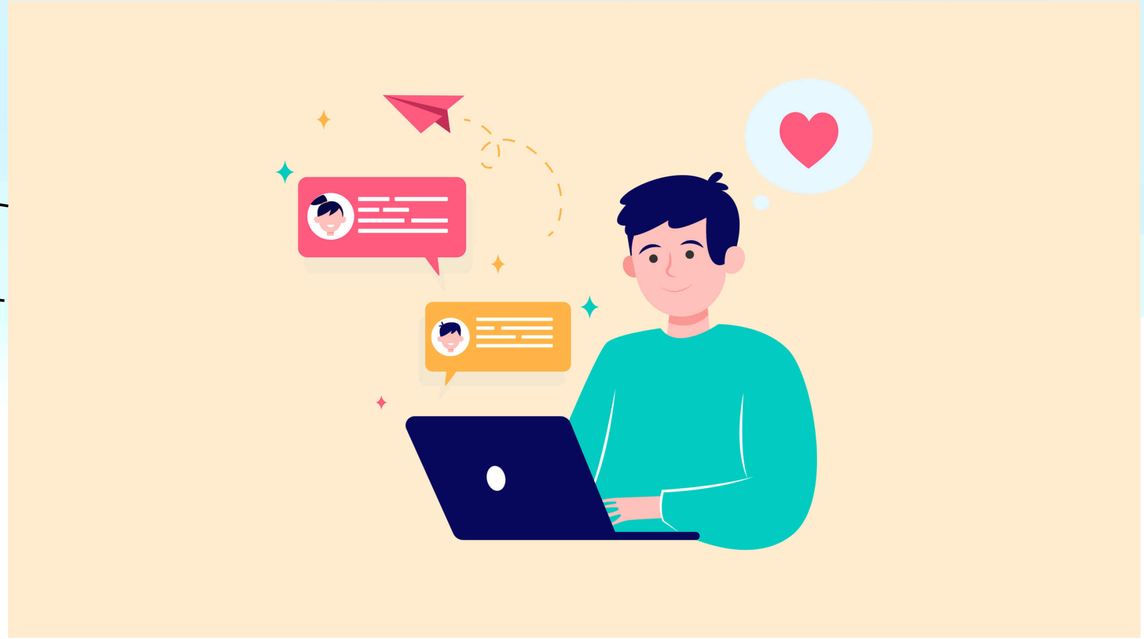
Todays’ Agenda

- What is Auto Scaling Group?
- High Availability and how to achieve it with ASG?
- Network Layers
- Fault Tolerance using Load-Balancer
- Automation using AMI

High Availability

Aims to Ensure an agreed level of operational performance, usually uptime, for a higher than normal period



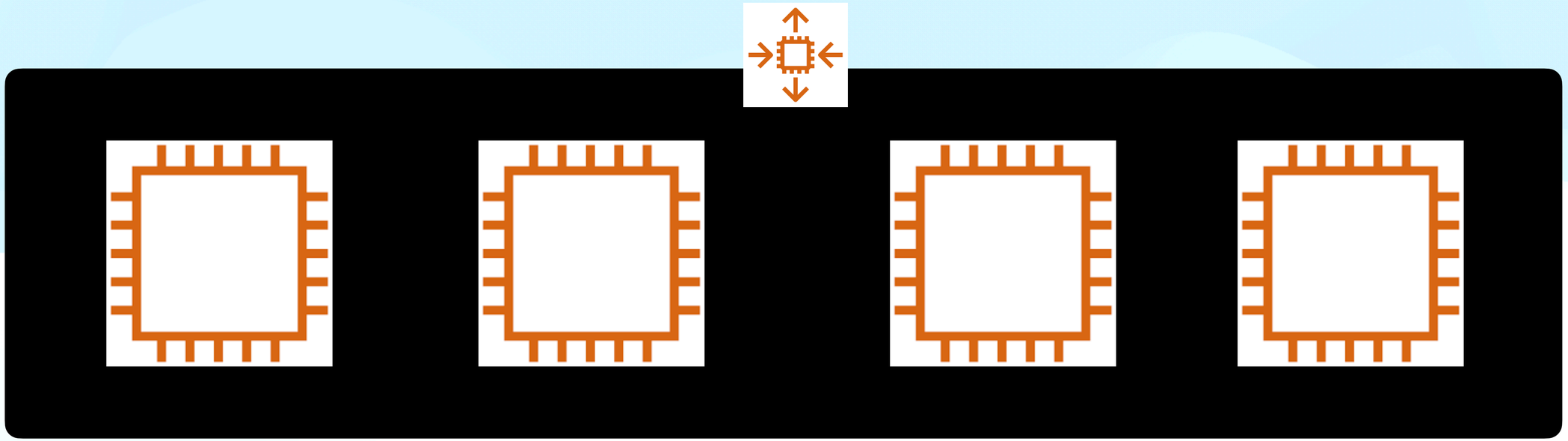


99.9 % Availability

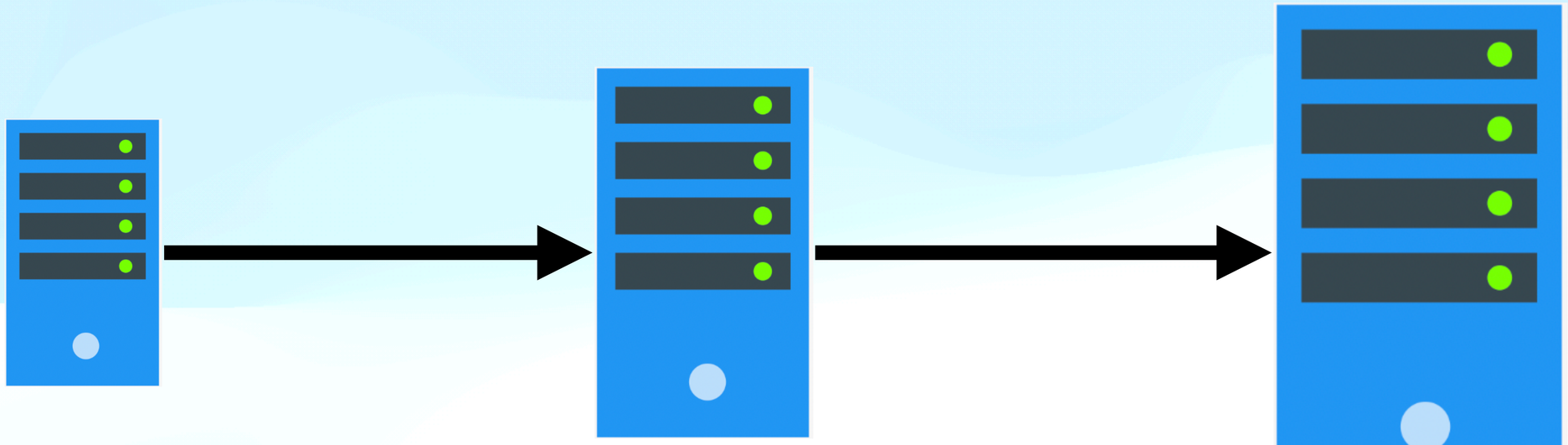
99.9999999 % Availability

AWS AutoScaling Groups

Launch Template
Provides EC2 Config



Vertical Scaling

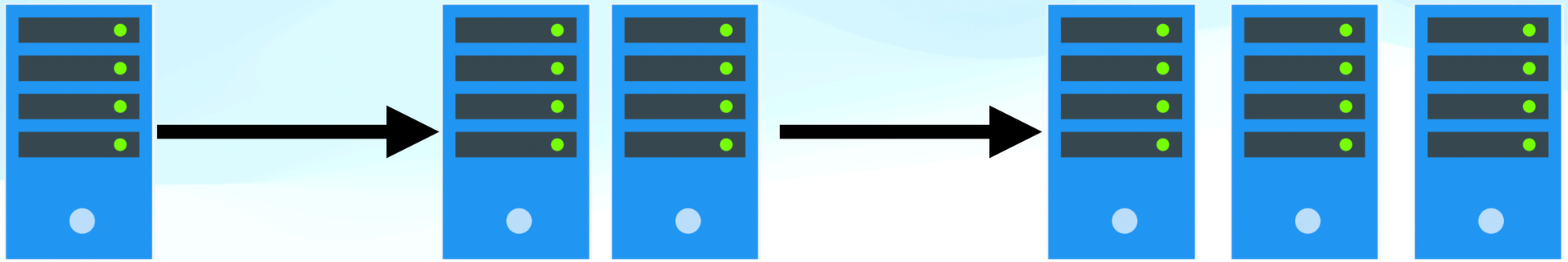


2 vCPU
8 GiB RAM

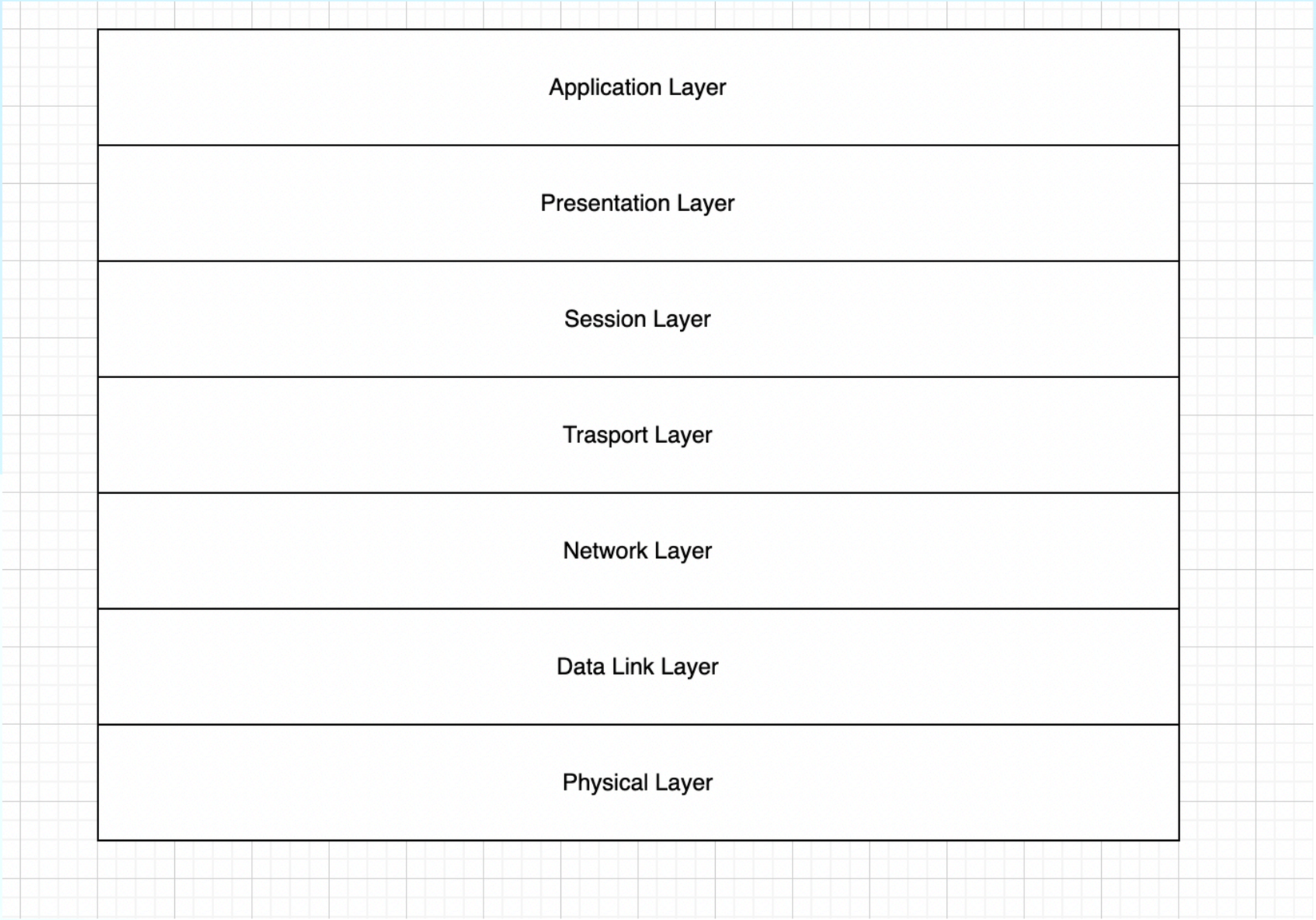
4 vCPU
16 GiB RAM

8 vCPU
32 GiB RAM

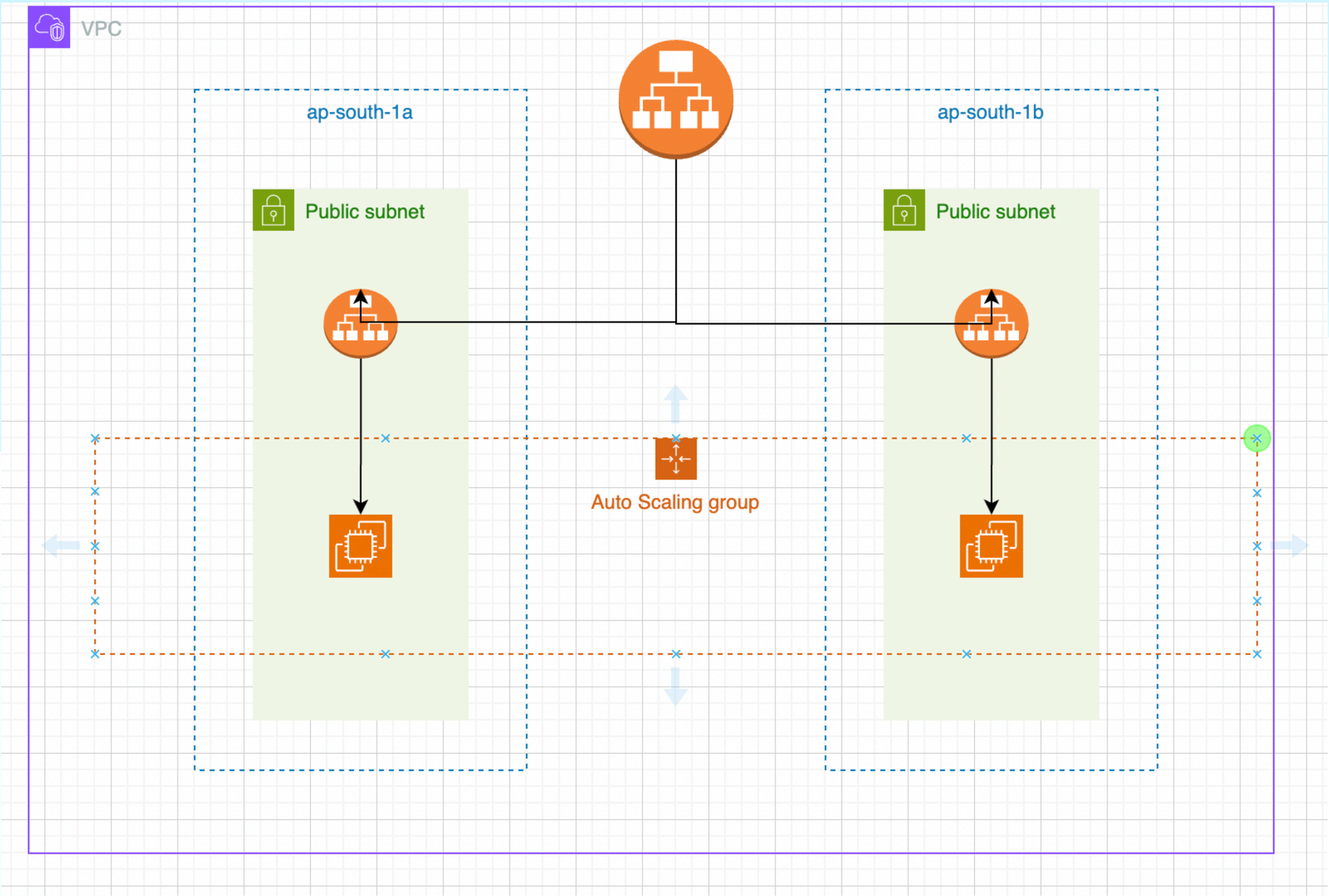
Horizontal Scaling



Network Layers



Load Balancers



Application Load Balancers

- Layer 7 Load Balancer, listens on HTTP or HTTPS
- Supports Layer 7 Path based routing, Headers, Cookies, User Location, etc
- ALBs or Slower than NLB
- Checks for the Application Health
- Listeners will have rules which will forward the traffic to Target Groups
- Can check for Application health checks

Network Load Balancers

- Layer 4 load balancer. Supports TCP, UDP, TLS, TCP_UDP
- No visibility or understanding of HTTP or HTTPS
- No path based routing, no headers and no cookies
- Very fast than ALB
- Just do the TCP Handshake, can’t do the application health checks.

QUIZ TIME

Q & A