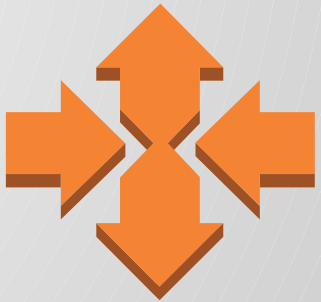
REA AWS Training

AutoScaling and ELBs

So you have an EC2 instance

- You created an instance
- SSH into it
- Installed your app and a web server
- It's serving traffic

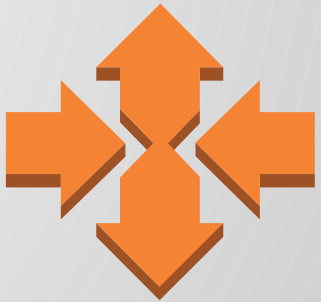
- What's if
  - The instance dies
  - Or it's too small...



Auto Scaling

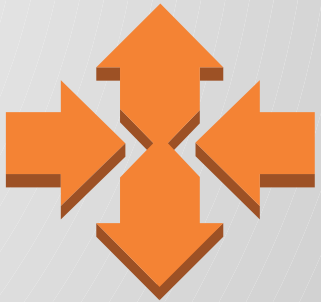
- Capacity on demand
- Resilience
- You provide parameters:
  - Min, Max, Desired capacities
- Scaling policies triggered manually or via CloudWatch
- Scheduled actions

- How do I automate my app deployment?



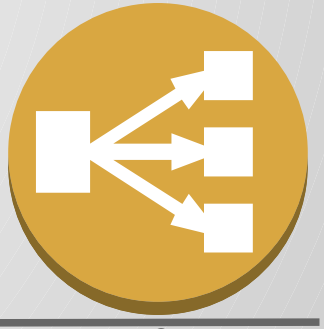
Launch Config

- Specified in the ASG
- Has everything your instance needs:
  - AMI ID
  - Instance Size
  - User Data script to deploy your app
    - Can use cloud-init or cfn-init for more advanced tasks
- Keep speed in mind
  - mitigations: baking AMI, minimal install



Elastic Load Balancing (ELB)

- Routes traffic to one or more back-end instances
- HTTP, HTTPS, or arbitrary TCP port
- Cross-Zone load balancing
- Health check validates back-end instance
- “scheme: internal” vs “scheme: internet-facing”
- Security groups protect ELBs



Autoscaling, ELB hands-on

- Using autoscaling to provide Multi-AZ resilience to instance termination or problems
- Using IAM to get access to another resource
- Using autoscaling to scale-out and back in as load changes
- Using autoscaling to schedule services up and down