AWS EC2

Elastic Compute Cloud

A service that provides secure, resizable compute capacity in the cloud

It is designed to make web-scale cloud computing easier

EC2 Provides:
- A virtual computing environment
- Preconfigured environments/images called AMI's
  - Allows you to quickly setup a virtual machine
- You can configure CPU, Memory, Storage, OS...
- Has secure login via ssh keys, and firewalls to prevent unwanted traffic

The EC2 has a firewall known as Security Groups
- Instance level
- Series of permissions associated with IP's
- Deny by default

Elastic Block Storage
- Extra storage for an EC2
- Setup and attach individual EC2's

EC2 AutoScaling

Allows you to scale up or down resources for your application automatically ensuring the availability of your application

Two ways to autoscale

Vertical: giving the server more resources
- CPU cores/Memory

Horizontal: spin up extra servers, and split the load between them

AWS uses horizontal scaling

Benefits of Autoscaling
- Improved fault tolerance
- Increased application availability
- Lower costs

EC2 Autoscaling Groups

Collections of EC2's with similar characteristics
- Use these to increase the number of EC2's
- Use these to decrease the number of EC2's
- Use these to set a fixed amount of EC2's

Launch configurations allow you set the parameters for autoscaling

Scheduled Scaling
Dynamic Scaling
Predictive Scaling