

# Spring - Mar 11

## Agenda

- 1, introduction to AWS
- 2, Compute in Cloud
- 3, Global infrastructure and Security
- 4, Networking
- 5, Storage and database
- 6, Security
- 7, Monitoring and Analytics
- 8, Migration

## 0, AWS Certification

foundational

associate

Professional

- solutions architect
- devops

Specialty

## 1, introduction to AWS

### client - server

server

- CPU
- RAM

- Data
- Routers, Switch

## **Traditional Approach**

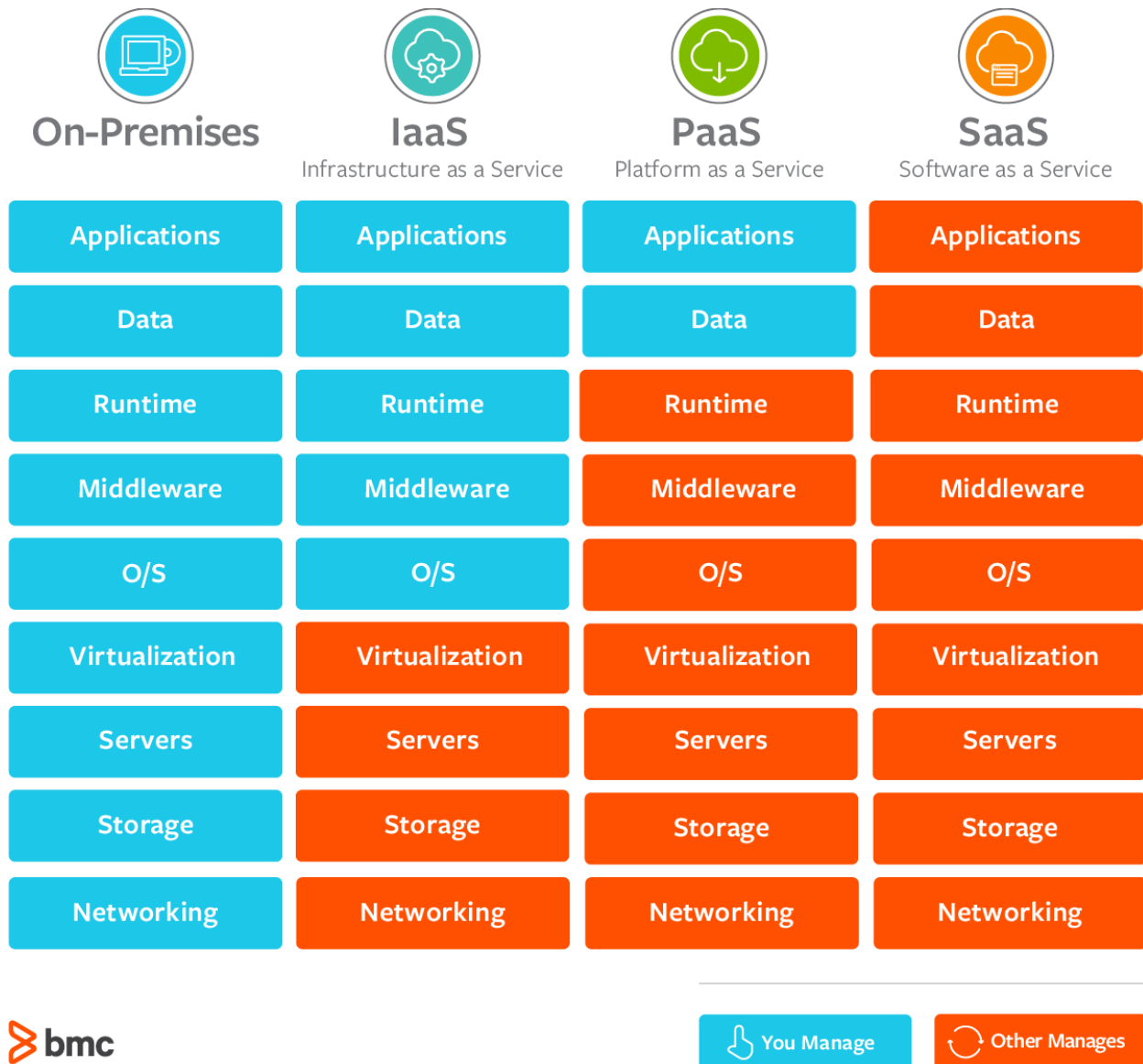
- pay for the rent for data center
- pay for power supply, cooling, maintenance
- adding and replacing hardware
- scaling is limited
- monitor 24/7

## **Cloud Computing**

*on demand delivery of IT resource*

- pay as you go pricing
  - how much pay depends on how much you use
- provision exactly the right size and type of computing resources
- accessed instantly

## **SaaS vs PaaS vs IaaS**



## Deployment models for cloud computing

- public cloud: cloud based
  - put everything on the cloud
- private cloud: on-premises
- Hybrid cloud: hybrid
  - some parts of your services are in the local, and some parts are in the cloud

# 2, Compute in Cloud

## EC2

### Amazon Elastic Compute Cloud

- general purpose instance
  - provide a very balanced compute memory and networking resources
  - application server/gaming server/backend server for enterprise application
- compute optimized instance
- memory optimized instance
  - process large data size in the memory
- accelerated computing instance
  - use a hardware accelerators and processors
    - double, treble cores of processors
- storage optimized instance
  - designed for workloads that require high sequential read and write access to large data sets on local storage.

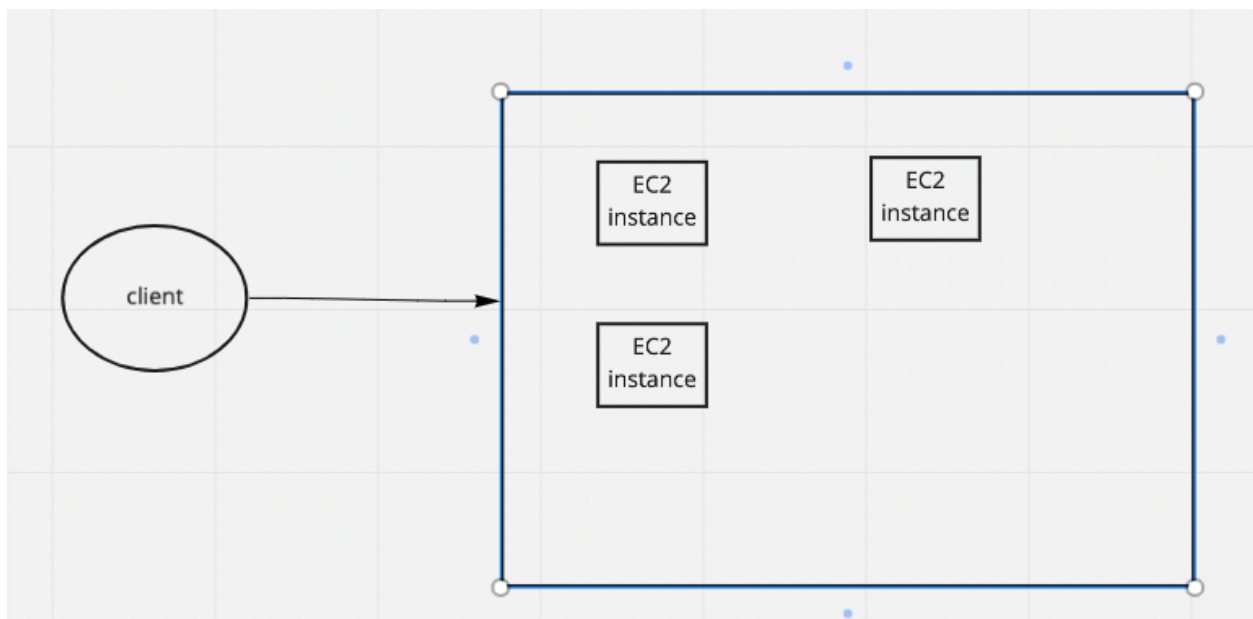
### EC2 Pricing

- spot instance
- reserved instance
- on demand instance

### Scaling

vertical scaling/horizontal scaling

- **Auto Scaling - AWS(important concept)**
  - belong to a kind of horizontal scaling (automatically)
  - price depends on how much you really use

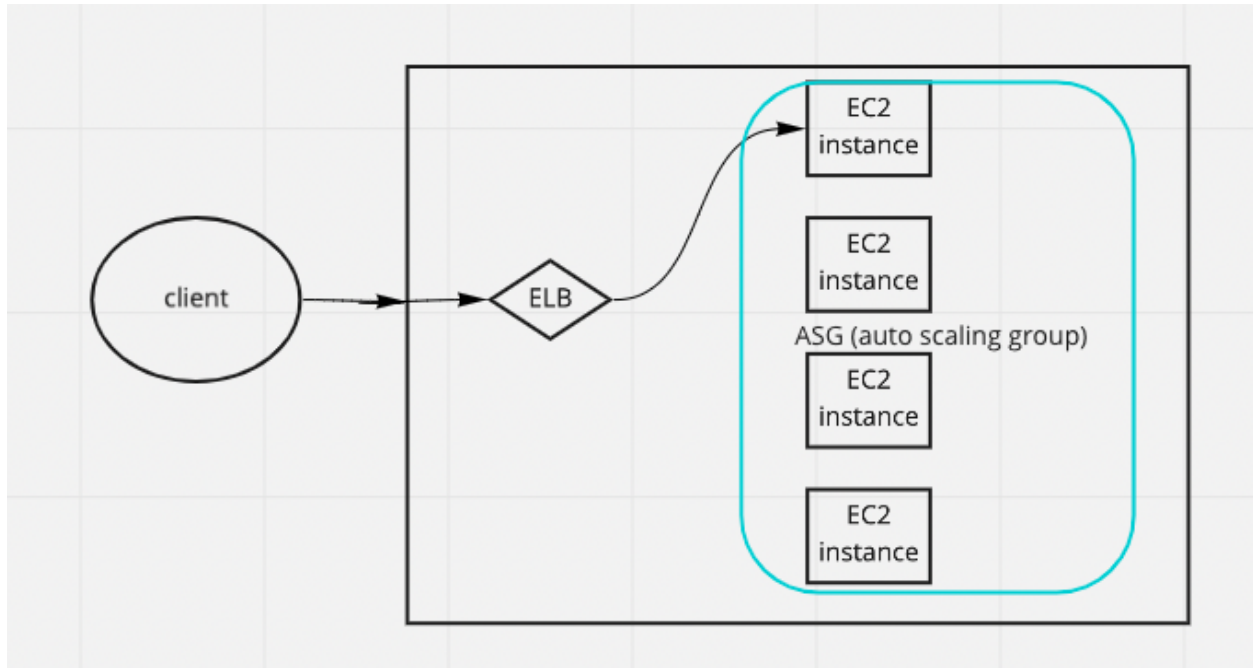


in this case

- originally, provide two instances for client
- when more requests come in, the Auto Scaling will add one more instance to fit the request requirements.
- if one is not enough, add more

- you just pay how much you use

## Elastic Load Balancing



In this case

- When request is sent by client, ELB will decide which EC2 instance to handle the request
- sticky session
  - route the specific request to specific session

ASG - auto scaling group

## Monolithic application vs Microservice Application

- Monolithic application
  - closely coupled: if one component is down, may the whole server will be down
  - easy to establish(e.g less than 100 users)
- MicroServices (very suitable for AWS)
  - loosely coupled

Attention:

One instance can service for several MicroService

One MicroService also may demand multiple instances to support

Instance and MicroService are not direct relationship

## **SNS (Amazon Simple Notification Service)**

## **SQS (Amazon Simple Queue Service)**

## **AWS lambda (serverless)**

## **Virtualization**

*Virtualization is a software based or just a representation of application server or storage and network.*

- virtual machine
- hypervisor
  - hypervisor is a software which will create a virtual machine for you.

## **Containers**

- docker

EC2 is virtualization based

## **ECS (Amazon Elastic Container Service)**

- container base
- support docker containers
  - just deploy docker in the ECS

## **ECR (Amazon Elastic Container Registry/Repository)**

- use to manage dockers

- inventory

## **EKS (Amazon Elastic Kubernetes Service)**

k8s

**Kubernetes**, also known as K8s, is an open-source system for automating deployment, scaling, and management of containerized applications.

- use to manage all (e.g 100)dockers?

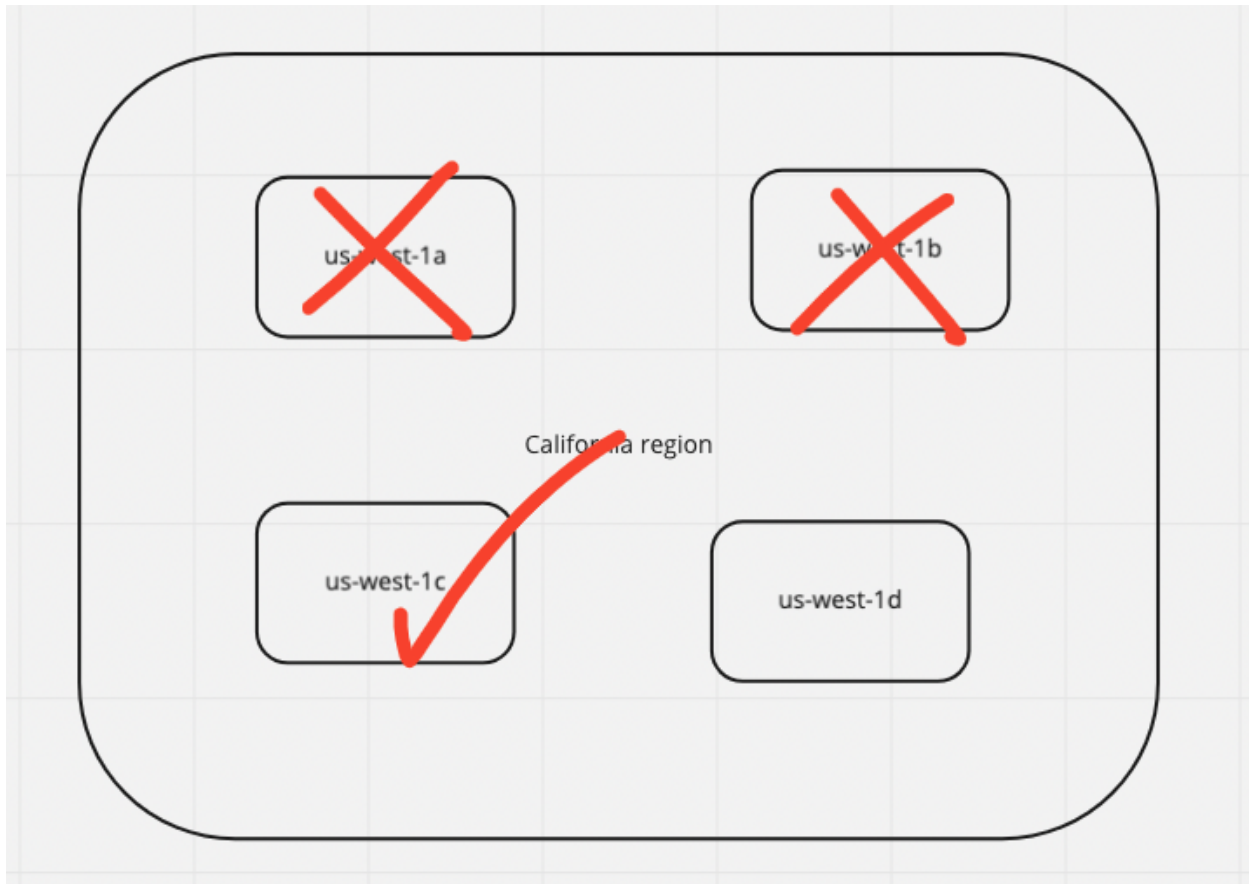
## **AWS Fargate**

- manage infrastructure

# **3, Global infrastructure and Security**

## **AWS Region & Availability Zone**





## Edge Location

- cache

## AWS Elastic Beanstalk

- adjust capacity
- load balancing
- automatic scaling
- application health monitoring

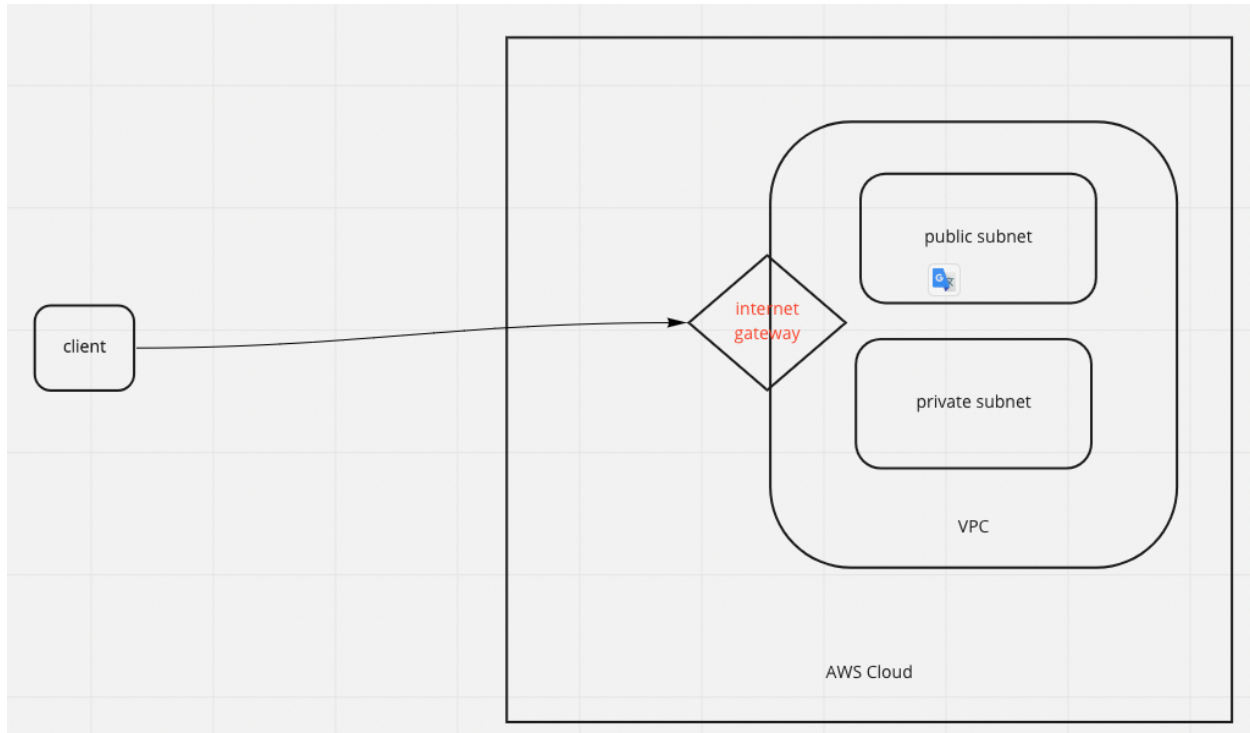
## AWS CloudFormation

## 4, Networking

## VPC (Amazon Virtual Private Cloud)

- public subnets
- private subnets

## Internet gateway



## Virtual private gateway

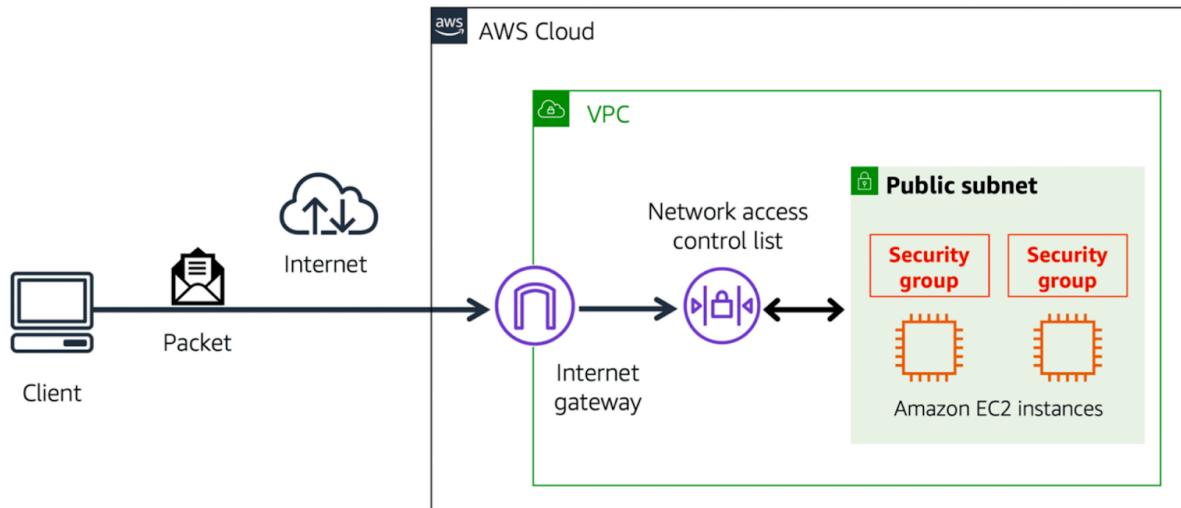
## AWS Direct Connect

## ACLs (Networking access control lists )

- control inbound and outbound traffic at the subnet level

## Security group

- control inbound and outbound traffic for an Amazon EC2 instance



## Route 53

- is a DNS web service.

# 5, Storage and database

## EBS (Elastic Block Store)

- incremental backup

## S3 (Simple Storage Service)

- object level storage

s3 storage classes

- s3 standard
- s3 standard- infrequent access
- s3 one zone - infrequent access
- s3 intelligent tiering
- s3 glacier
- s3 glacier deep archive

# Amazon Elastic File storage

## RDS

- relational database
- database engines
  - Aurora
  - PostgreSQL
  - MySQL
  - MariaDB
  - Oracle Database
  - SQL Server

## DynamoDB

## Redshift

## ElasticCache

- redis
- memocache

# 6, Security

shared responsibility model

- customer responsibility (security in the cloud)
- AWS responsibility (security of the cloud)

## IAM

- IAM users, group, roles
- IAM policies

## **AWS Organizations**

# **7, Monitoring and Analytics**

## **Amazon CloudWatch**

- metrics
- dashboard

## **AWS CloudTrail**

- record API calls for your account

# **8, Migration**

## **perspectives**

- business
- people
- governance
- platform
- Security
- Operations

## **physical devices**

- AWS Snowcone
- AWS Snowball
- AWS Snowmobile