AWS

What’s AWS?

AWS (Amazon Web Services) is a Cloud Provider

• They provide you with servers and services that you can use on demand and scale easily

Types of cloud computing

Infrastructure as a Service (IaaS)

• Provide building blocks for cloud IT

• Provides networking, computers, data storage space

• Highest level of flexibility

• Easy parallel with traditional on-premises IT

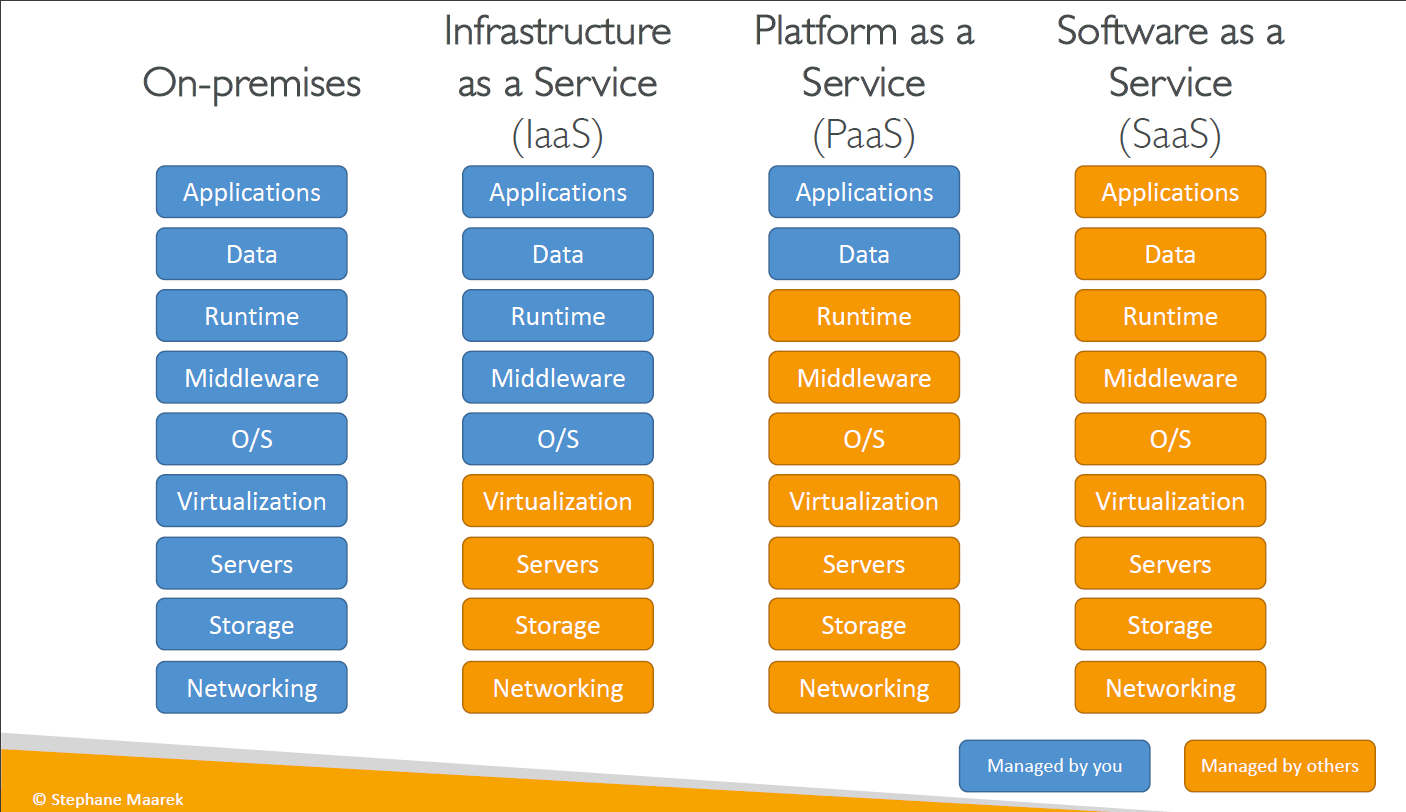
Platform as a Service (PaaS)

• Removes the need for your organization to manage the underlying infrastructure

• Focus on the deployment and management of your applications

Software as a Service (SaaS)

• Completed product that is run and managed by the service provider



AWS Regions

AWS has Regions all around the world

Names can be us-east-1, eu-west-3…, A region is a cluster of data centers. Most AWS services are region-scoped

• AWS Availability Zones

Each region has many availability zones (usually 3, min is 2, max is 6). Example:

• ap-southeast-2a

• ap-southeast-2b

• ap-southeast-2c

Each availability zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity

They’re separate from each other, so that they’re isolated from disasters

• AWS Data Centers

• AWS Edge Locations / Points of Presence

AWS Points of Presence (Edge Locations)

Amazon has 216 Points of Presence (205 Edge Locations & 11 Regional Caches) in 84 cities across 42 countries,Content is delivered to end users with lower latency

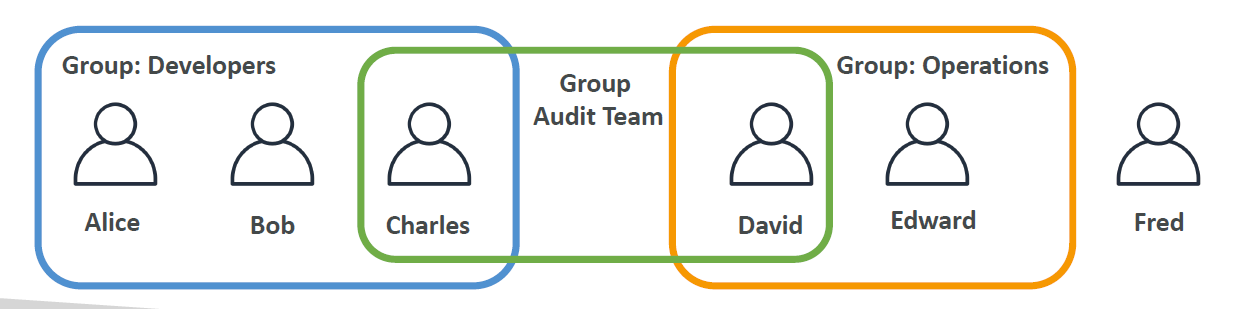
IAM (IDENTITY AND ACCESS MANAGEMENT)

Root account created by default, shouldn’t be used or shared

• Users are people within your organization, and can be grouped

• Groups only contain users, not other groups

• Users don’t have to belong to a group, and user can belong to multiple groups



Policies

Users or Groups can be assigned JSON documents called policies

• These policies define the permissions of the users

• In AWS you apply the least privilege principle: don’t give more permissions than a user needs



Policies structure

Consists of

Version: policy language version, always include “2012-10-

17”

• Id: an identifier for the policy (optional)

• Statement: one or more individual statements (required)

• Statements consists of

• Sid: an identifier for the statement (optional)

• Effect: whether the statement allows or denies access (Allow, Deny)

• Principal: account/user/role to which this policy applied to

• Action: list of actions this policy allows or denies

• Resource: list of resources to which the actions applied to

• Condition: conditions for when this policy is in effect (optional)

AWS CLI

A tool that enables you to interact with AWS services using commands in your command-line shell

• Direct access to the public APIs of AWS services

• You can develop scripts to manage your resources

• It’s open-source https://github.com/aws/aws-cli

• Alternative to using AWS Management Console

AWS Software Development Kit (AWS SDK)

• Language-specific APIs (set of libraries)

• Enables you to access and manage AWS services programmatically

• Embedded within your application

• Supports

• SDKs (JavaScript, Python, PHP, .NET, Ruby, Java, Go, Node.js,C++)

• Mobile SDKs (Android, iOS, …)

• IoT Device SDKs (Embedded C, Arduino, …)

• Example: AWS CLI is built on AWS SDK for Python

IAM Roles for Services

• Some AWS service will need to perform actions on your behalf

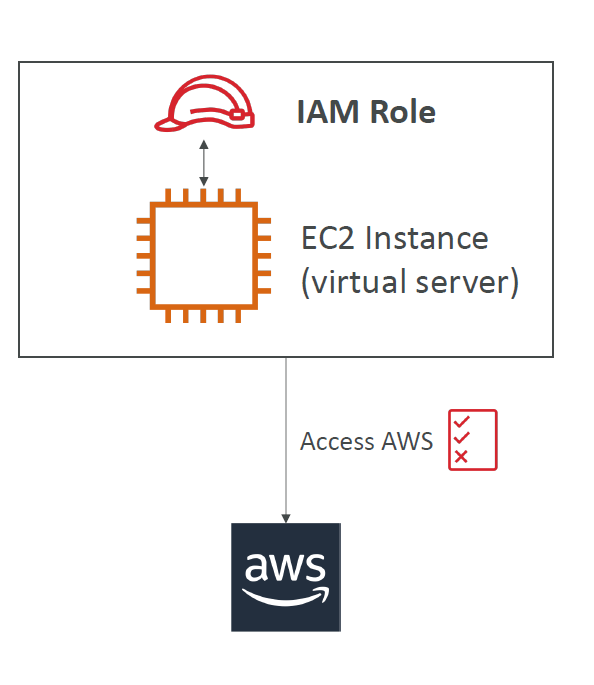
• To do so, we will assign permissions to AWS services with IAM Roles

Common roles:

• EC2 Instance Roles

• Lambda Function Roles

• Roles for CloudFormation



IAM Security Tools

IAM Credentials Report (account-level)

• a report that lists all your account's users and the status of their various credentials

IAM Access Advisor (user-level)

• Access advisor shows the service permissions granted to a user and when those sservices were last accessed.

• You can use this information to revise your policies.



EC2 (ELASTIC CLOUD COMPUTE)

EC2 mainly gives us the capability of:

1. Renting virtual machines (EC2)
2. Storing data on virtual drives (EBS)
3. Distributing load across machines (ELB)
4. Scaling the services using an auto-scaling group (ASG)

EC2 Instance types

1. General Purpose

Great for a diversity of workloads such as web servers or code repositories. It provides balance between compute, memory and networking.

1. Compute Optimized

Great for compute-intensive tasks that require high performance processors:

Batch processing workloads, Media transcoding, High performance web servers, High performance computing (HPC), Scientific modeling & machine learning, Dedicated gaming servers

1. Memory Optimized

Fast performance for workloads that process large data sets in memory

1. Storage Optimized

Great for storage-intensive tasks that require high, sequential read and write

access to large data sets on local storage.

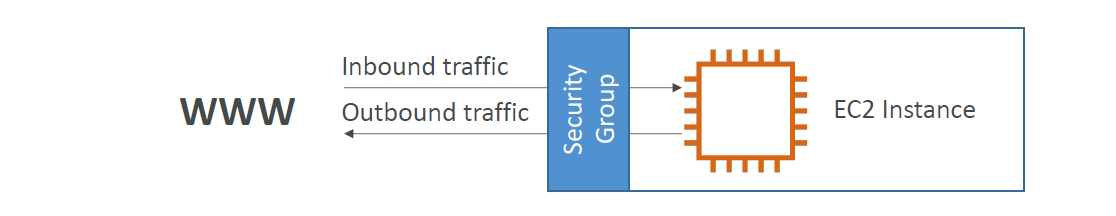
SECURITY GROUPS

Security Groups are the fundamental of network security in AWS

They control how traffic is allowed into or out of our EC2 Instances

Security groups only contain ALLOW rules

Security groups rules can reference by IP or by security group



Security groups are acting as a “firewall” on EC2 instances

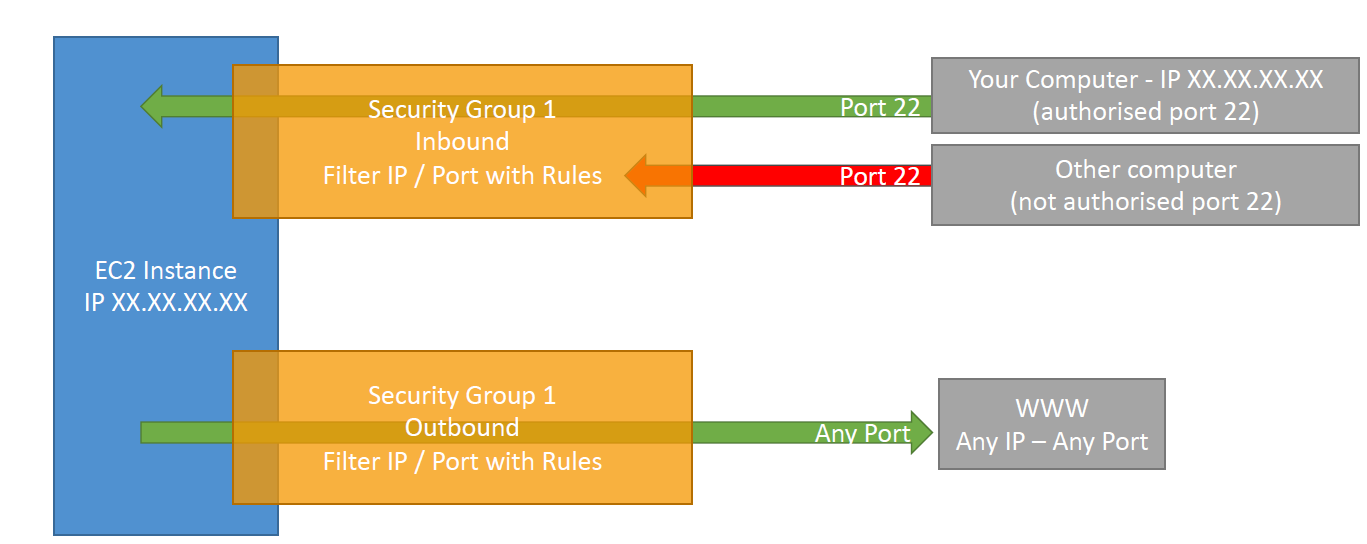
They regulate:

• Access to Ports

• Authorised IP ranges – IPv4 and IPv6

• Control of inbound network (from other to the instance)

• Control of outbound network (from the instance to other)

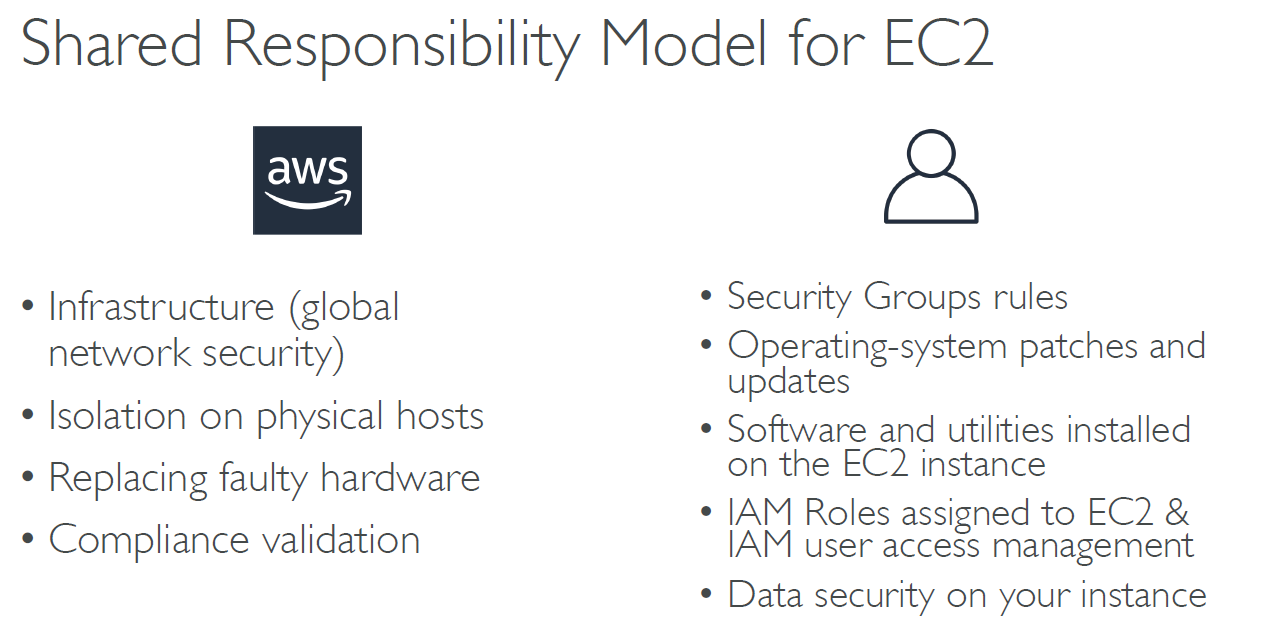


EC2 Instances Purchasing Options

1. On-Demand Instances – short workload, predictable pricing, pay by second
2. Reserved (1 & 3 years)

* Reserved Instances – long workloads
* Convertible Reserved Instances – long workloads with flexible instances

1. Savings Plans (1 & 3 years) –commitment to an amount of usage, long workload
2. Spot Instances – short workloads, cheap, can lose instances (less reliable)
3. Dedicated Hosts – book an entire physical server, control instance placement
4. Dedicated Instances – no other customers will share your hardware
5. Capacity Reservations – reserve capacity in a specific AZ for any duration



EBS( ELASTIC BLOCK STORAGE)

* An EBS (Elastic Block Store) Volume is a network drive you can attach to your instances while they run.
* It allows your instances to persist data, even after their termination

Features

1.It’s a network drive (i.e. not a physical drive)

•It uses the network to communicate the instance, which means there might be a bit of latency

• It can be detached from an EC2 instance and attached to another one quickly

2.It’s locked to an Availability Zone (AZ)

• An EBS Volume in us-east-1a cannot be attached to us-east-1b

• To move a volume across, you first need to snapshot it

3.Have a provisioned capacity (size in GBs, and IOPS)

• You get billed for all the provisioned capacity

• You can increase the capacity of the drive over time

Delete on termination attribute

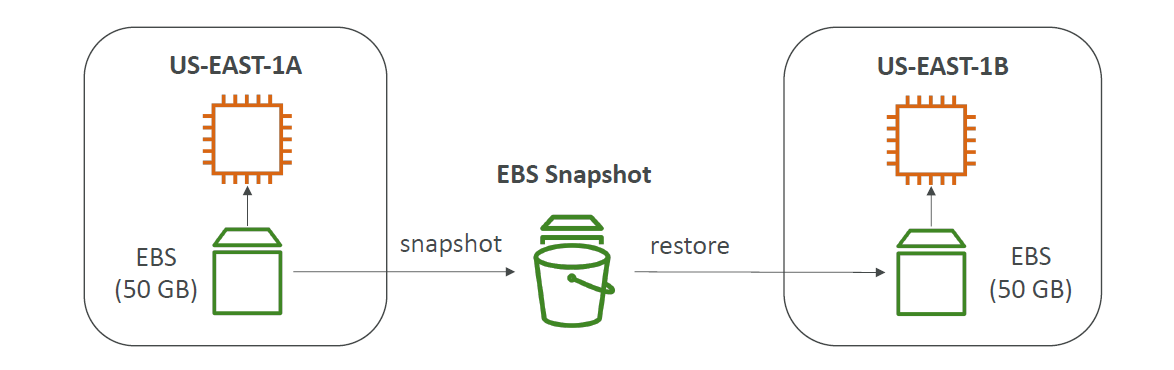
Controls the EBS behaviour when an EC2 instance terminates

• By default, the root EBS volume is deleted (attribute enabled)

• By default, any other attached EBS volume is not deleted (attribute disabled)

Snapshots

* Make a backup (snapshot) of your EBS volume at a point in time
* Not necessary to detach volume to do snapshot, but recommended
* Can copy snapshots across AZ or Region



SNAPSHOTS Features:

1. Snapshots Archive

* Move a Snapshot to an ”archive tier” that is 75% cheaper
* Takes within 24 to 72 hours for restoring the archive

1. Recycle bin for snapshots

* Setup rules to retain deleted snapshots so you can recover them after an accidental deletion
* Specify retention (from 1 day to 1 year)

AMI(AMAZON MACHINE IMAGE)

AMI IS a customization of an EC2 instance

• You add your own software, configuration, operating system, monitoring…

• Faster boot / configuration time because all your software is pre-packaged

You can launch EC2 instances from:

• A Public AMI: AWS provided

• Your own AMI: you make and maintain them yourself

• An AWS Marketplace AMI: an AMI someone else made (and potentially sells)

EC2 IMAGE BUILDER

Used to automate the creation of Virtual Machines or container images. Automate the creation, maintain, validate and test EC2 AMIs.

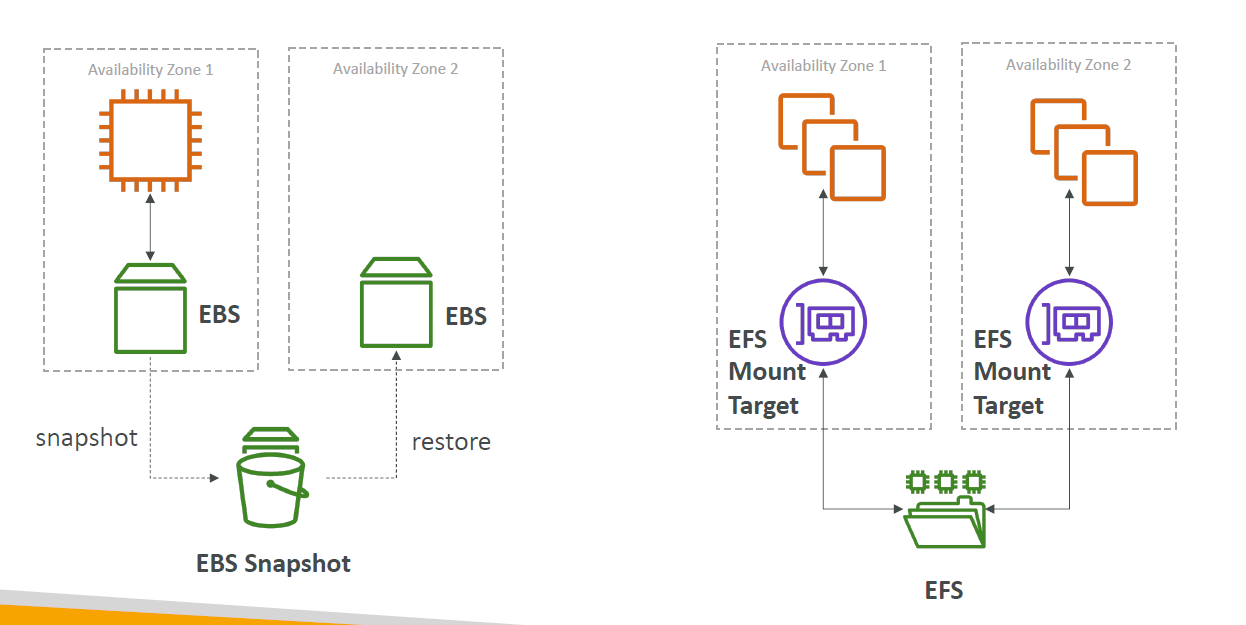
EC2 INSTANCE STORE

EBS volumes are network drives with good but “limited” performance. If you need a high-performance hardware disk, use EC2 Instance Store.

* Better I/O performance
* EC2 Instance Store lose their storage if they’re stopped (ephemeral)
* Good for buffer / cache / scratch data / temporary content
* Risk of data loss if hardware fails
* Backups and Replication are your responsibility

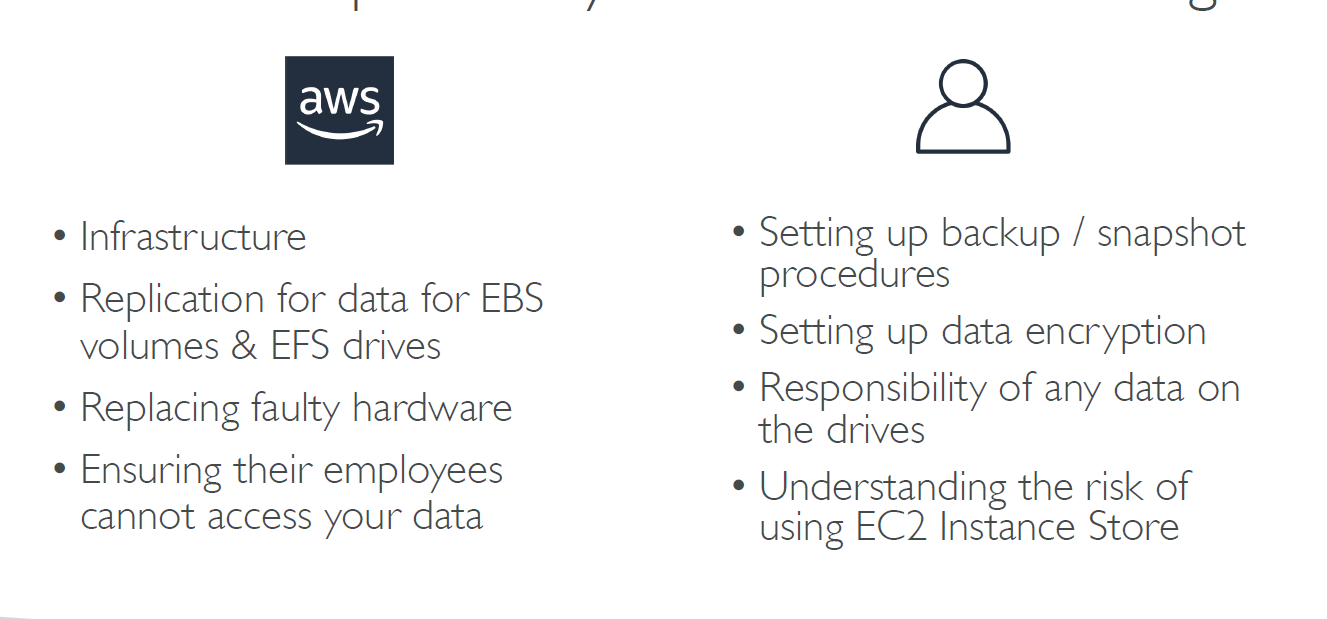
ELASTIC FILE SYSTEM

Managed NFS (network file system) that can be mounted on 100s of EC2



EFS – IA (INFREQUENT ACCESS)

Storage class that is cost-optimized for files not accessed every day. EFS will automatically move your files to EFS-IA based on the last time they were accessed



AMAZON FSX

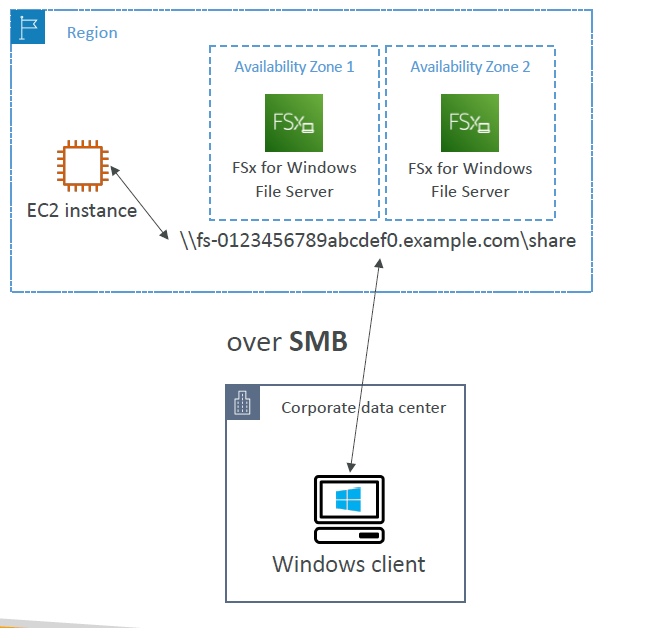
Launch 3rd party high-performance file systems on AWS

1. Windows File Server

A fully managed, highly reliable, and scalable Windows native shared

file system. Built on Windows File Server.

It supports SMB protocol.



1. Lustre:

A fully managed, high-performance, scalable file storage for High Performance Computing (HPC)

ELASTIC LOAD BALANCING

Scalability means that an application / system can handle greater loads by adapting.

There are two kinds of scalability:

• Vertical Scalability

Vertical Scalability means increasing the size of the instance

• Horizontal Scalability (= elasticity)

Horizontal Scalability means increasing the number of instances / systems for your application

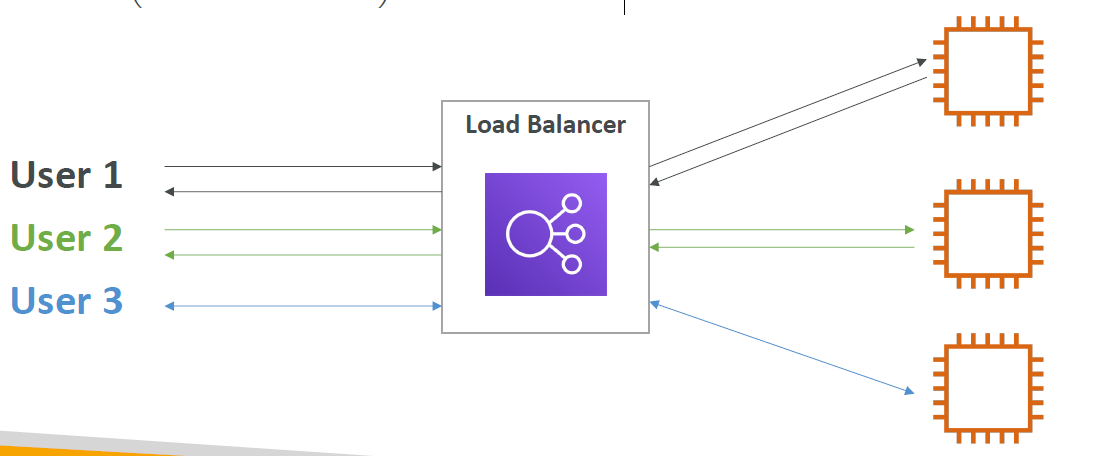
High availability means running your application / system in at least 2 Availability Zones.

Scalability: ability to accommodate a larger load by making the hardware stronger (scale up), or by adding nodes (scale out).

Elasticity: once a system is scalable, elasticity means that there will be some “auto-scaling” so that the system can scale based on the load. This is “cloud-friendly”: pay-per-use, match demand, optimize costs.

Agility: (not related to scalability - distractor) new IT resources are only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes.

Load balancers are servers that forward internet traffic to multiple servers (EC2 Instances) downstream.



AUTO SCALING GROUP

In real-life, the load on your websites and application can change

In the cloud, you can create and get rid of servers very quickly

The goal of an Auto Scaling Group (ASG) is to:

• Scale out (add EC2 instances) to match an increased load

• Scale in (remove EC2 instances) to match a decreased load

• Ensure we have a minimum and a maximum number of machines running

• Automatically register new instances to a load balancer

• Replace unhealthy instances

Scaling strategies:

1. Manual Scaling

2. Dynamic Scaling

* Simple / Step Scaling
* Target Tracking Scaling
* Scheduled Scaling

3. Predictive Scaling

S3 (AMAZON SIMPLE STORAGE SERVICE)

Amazon S3 allows people to store objects (files) in “buckets” (directories)

Buckets must have a globally unique name (across all regions all accounts)

Websites

S3 can host static websites and have them accessible on the www.

Versioning

You can version your files in Amazon S3

• It is enabled at the bucket level

• Same key overwrite will increment the “version”: 1, 2, 3….

S3 Access logs

For audit purpose, you may want to log all access to S3 buckets

Any request made to S3, from any account, authorized or denied, will be logged into another S3 bucket

Replication

1.Cross Region Replication:

2. Same Region Replication: