Computing Services

Elastic Compute Cloud (EC2) allows you to launch **Virtual Machines (VM)**

What is a Virtual Machine?
A Virtual Machine (VM) is an emulation of a physical computer using software.
Server Virtualization allows you to easily **create**, **copy**, **resize** or **migrate** your server.
Multiple VMs can run **on the same physical server** so you can share the cost with other customers.
Imagine if your server or computer was an executable file on your computer

When we launch a Virtual Machine we call it an "**instance**"

EC2 is **highly configurable server** where you can choose **AMI** that affects options such as:

- The amount of CPUs
- The amount of Memory (RAM)
- The amount of Network Bandwidth
- The Operation System (OS) eg. Windows 10, Ubuntu, Amazon Linux 2
- Attach multiple virtual hard-drives for storage eg. Elastic Block Store (EBS)

An Amazon Machine Image (AMI) is a predefined configuration for a Virtual Machine.

EC2 is also considered **the backbone of AWS** because the majority of AWS services are using EC2 as their underlying servers. eg. S3, RDS, DynamoDB, Lambdas

Computing Services

Virtual Machines — an emulation of a physical computer using software

Amazon LightSail is the **managed virtual server service**. It is the "friendly" version of EC2 Virtual Machines
When you need to launch a Linux or Windows server but don't have much AWS knowledge. eg. Launch a Wordpress

Containers — virtualizing an Operation System (OS) to run multiple workloads on a single OS instance. Containers are generally used in micro-service architecture (when you divide your application into smaller applications that talk to each other)

Elastic Container Service (ECS) is a **container orchestration service** that support **Docker** containers. Launches a cluster of server(s) on EC2 instances with Docker installed. ***When you need Docker as a Service, or you need to run containers.***

Elastic Container Registry (ECR) is **repository for container images**. In order to launch a containers you need an image. An image just means a saved copy. A repository just means a storage that has version control.

ECS Fargate is **serverless orchestration container service**. It is the same as ECS expect you pay-on-demand per running container (With ECS you have to keep a EC2 server running even if you have no containers running) AWS manages the underlying server, so you don't have to scale or upgrade the EC2 server.

Elastic Kubernetes Service (EKS) is **a fully managed Kubernetes service**. Kubernetes (K8) is an open-source orchestration software that was created by Google and is generally the standard for managing microservices. ***When you need to run Kubernetes as a Service.***

Serverless — when the underlying servers are managed by AWS. You don't worry or configure servers.

AWS Lambda is a **serverless functions service.** You can run code without provisioning or managing servers. You upload small pieces of code, choose much memory and how long function is allowed to run before timing out. You a charged based on the runtime of the serverless function rounded to the nearest 100ms.

Higher Performance Computing Services

The Nitro System A combination of **dedicated hardware and lightweight hypervisor** enabling faster innovation and enhanced security. All new EC2 instance types use the Nitro System.

- Nitro Cards — specialized cards for VPC, EBS and Instance Storage and controller card
- Nitro Security Chips — Integrated into motherboard. Protects hardware resources.
- Nitro Hypervisor — lightweight hypervisor Memory and CPU allocation Bare Metal-like performance

Bare Metal Instance You can launch EC2 instance that have no hypervisor so you can run workloads directly on the hardware for maximum performance and control. The **M5 and R5** EC2 instances run are bare metal.

Bottlerocket is a Linux-based open-source operation system that is purpose-built by AWS for running containers on Virtual Machines or bare metal hosts

What is High Performance Computing (HPC)?

A cluster of hundreds of thousands of servers with fast connections between each of them with the purpose of boosting computing capacity.

When you need a supercomputer to perform computational problems too large to run on a standard computers or would take to long.

AWS ParallelCluster is an **AWS-supported open source cluster management tool** that makes it easy for you to deploy and manage High Performance Computing (HPC) clusters on AWS.

Edge and Hybrid Computing Services

What is Edge Computing?

When you push your computing workloads outside of your networks to run close to the destination location.
eg. Pushing computing to run on phones, IoT Devices, or external servers not within your cloud network.

What is Hybrid Computing?

When you're able to run workloads on both your on-premise datacenter and AWS Virtual Private Cloud (VPC)

AWS Outposts is **physical rack of servers** that you can put in your data center. AWS Outposts allows you to use AWS API and Services such as EC2 right in your datacenter.

verizon vodafone business

AWS Wavelength allows you **to build and launch your applications in a telecom datacenter**. By doing this your applications with have ultra-low latency since they will be pushed over a the **5G network** and be closest as possible to the end user.

VMWare Cloud on AWS allows you to **manage on-premise virtual machines using VMWare** as EC2 instances. The data-center must being using VMWare for Virtualization.

vmware vSphere

AWS Local Zones are **edge datacenters located outside of an AWS region** so you can use AWS closer to end destination.

When you need faster computing, storage and databases in populated areas that are outside of an AWS Region

Cost and Capacity Management Computing Services

Cost Management How do we save money?
Capacity Management How do we meet the demand of traffic and usages though adding or upgrading servers?

EC2 Spot Instances, Reserved Instanced and Savings Plan
Ways to save on computing, by paying up in full or partially, by committing to a yearly contracts or by being flexible about availability and interruption to computing service.

AWS Batch plans, schedules, and executes **your batch computing workloads** across the full range of AWS compute services, can utilize Spot Instance to save money.

AWS Compute Optimizer suggests how to **reduce costs and improve performance** by using machine learning to analyze you previous usage history

EC2 Autoscaling Groups (ASGs)
Automatically adds or remove EC2 servers to meet the current demand of traffic. Will save you money and meet capacity since you only run the amount of servers you need.

Elastic Load Balancer (ELB)
Distributes traffic to multiple instance, can re-route traffic from unhealthy instance to healthy instances. can route traffic to EC2 instances running in different Availability Zones

AWS Elastic Beanstalk (EB) is for easily deploying web-applications without developers having to worry about setting up and understanding the underlying AWS Services. Similar to **Heroku**.