

## Elastic Compute Cloud (EC2)

### >EC2 Overview

This is the most fundamental service in AWS. This is a MUST since it is in and out of the exam. This is a secure, resizable compute capacity in the cloud. Like a VM, only hosted in AWS instead of your own data center. Gives you the capacity you want when you need it. You're in complete control over it.

EC2 allows you to rent and manage virtual servers in the cloud.

- Elastic compute power
  - You can stretch an elastic band far beyond its resting state. But part of what makes it truly elastic is the fact that, when you let go of it, it immediately returns to its original size. The reason the word elastic is used in the names of so many AWS services (Elastic Compute Cloud, Elastic Load Balancing, Elastic Beanstalk, and so on) is because those services are built to be easily and automatically resized.
  - Elastic load balancing
    - Elastic Load Balancing is the AWS service that automatically distributes incoming application traffic across multiple resources, such as Amazon EC2 instances.
    - Think of it as a balancer that “slows” but doesn't really slow down traffic.



The key is choosing the right tool for the right job. As traffic grows, the ELB grows/scales it is designed for the throughput.

A load balancer acts as a single point of contact for all incoming web traffic to your Auto Scaling group. This means that as you add or remove Amazon EC2 instances in response to the amount of incoming traffic, these requests route to the load balancer first. Then, the requests spread across multiple resources that will handle them. For example, if you have multiple Amazon EC2 instances, Elastic Load Balancing distributes the workload across the multiple instances so that no single instance has to carry the bulk of it.

- A load balancer distributes incoming application traffic across multiple EC2 instances in multiple Availability Zones. This increases the fault tolerance of your applications. Elastic Load Balancing detects unhealthy instances and routes traffic only to healthy instances.

Servers are physical compute hardware running in a data center. Which EC2 instances are the virtual servers running on these physical servers. Instances are not considered serverless. Because they actually exist on servers somewhere in the data center.

With over 500 instances and choice of the latest processor, storage, networking, operating system, and purchase model to help you best match the needs of your workload. We are the first major cloud provider that supports Intel, AMD, and Arm processors, the only cloud with on-demand EC2 Mac instances, and the only cloud with 400 Gbps Ethernet networking.

### EC2 Real World

Deploy a database to EX2 gives you full control over the database. Whereas deploying a web application allows multiple AZs to make the web application highly available.

You can access this through AWS management console, SSH, EC2 Instance connect, AWS Systems Manager

- AWS management console
  - You're able to configure and manage your instances via web browser.
- SSH
  - SSH allows a secure connection
- EC2 Instance connect
  - EIC allows you to use IAM policies to control SSH access to your instances, removing the need to manage SSH keys
- AWS Systems Manager
  - Allows you to manage your EC2 instances via a web browser or AWS CLI

The most common way to connect to Linux EC2 instances is via SSH. which the first thing you'll do is generate a key pair which consists of a private key and a public key. Proves your identity when connecting to an EC2

### > EC2 Pricing options

Several pricing options to choose from your EC2 instances

- On-Demand

## On-Demand

✓ A fixed price in which you are billed **down to the second** based on the instance type. There is no contract, and you pay only for what you use.

✓ **Use On-Demand instances when:**

- 1 You care about low cost without any upfront payment or long-term commitment.
- 2 Your applications have unpredictable workloads that **can't** be interrupted.
- 3 Your applications are under development.
- 4 Your workloads will **not** run longer than a year.

✓ **Fun facts:**

You can **reserve capacity** using On-Demand Capacity Reservations. The EC2 capacity is held for you whether or not you run the instance.

- Spot

## Spot

✓ Spot instances let you take advantage of **unused** EC2 capacity. Your request is fulfilled **only** if capacity is available.

✓ **Use Spot instances when:**

- 1 You are not concerned about the **start** or **stop** time of your application.
- 2 Your workloads **can** be interrupted.
- 3 Your application is only feasible at very low compute prices.

✓ **Fun facts:**

- 1 You can save up to **90%** off On-Demand prices.
- 2 You pay the spot price that's in effect at the beginning of each hour.

**Cheapest  
Option**

- Reserved Instances (RIs)

## Reserved Instances (RIs)

- ✓ **RIs** allow you to commit to a specific instance type in a particular Region for **1** or **3** years.
- ✓ **Use Reserved Instances when:**
  - 1 Your application has **steady state usage**, and you can commit to **1** or **3** years.
  - 2 You can pay money **upfront** in order to receive a discount on On-Demand prices.
  - 3 Your application requires a **capacity reservation**.
- ✓ **Fun facts:**
  - 1 You can save up to **75%** off On-Demand prices.
  - 2 You are required to sign a contract.
  - 3 You can **reserve capacity** in an **Availability Zone** for any duration.
  - 4 You can pay **All Upfront**, **Partial Upfront**, or **No Upfront**. **All Upfront** for the max term earns the highest discount.
  - 5 Provides **convertible types** at **54%** discount.

- - Dedicated Hosts

## Dedicated Hosts

- ✓ **Dedicated Hosts** allow you to pay for a physical server that is fully dedicated to running your instances.
- ✓ **Use Dedicated Hosts when:**
  - 1 You want to **bring your own** server-bound software **license** from vendors like Microsoft or Oracle.
  - 2 You have regulatory or corporate compliance requirements around tenancy model.
- ✓ **Fun facts:**
  - 1 You can save up to **70%** off On-Demand prices.
  - 2 You bring your existing **per-socket**, **per-core**, or **per-VM** software licenses.
  - 3 There is no multi-tenancy, meaning the server is not shared with other customers.
  - 4 A Dedicated Host is a physical server, whereas a Dedicated Instance runs on the host.

- - Savings plans

## Savings Plan

✓ **Savings Plan** allows you to commit to compute usage (measured per hour) for **1** or **3** years.

✓ **Use Savings Plans when:**

- 1 You want to lower your bill across multiple compute services.
- 2 You want the flexibility to change compute services, instance types, operating systems, or Regions.

✓ **Fun facts:**

- 1 You can save up to **72%** off On-Demand prices.
- 2 You are not making a commitment to a Dedicated Host, just compute usage.
- 3 Savings can be shared across various compute services like EC2, Fargate, and Lambda.
- 4 This does **not** provide a capacity reservation.

○

### >Using Roles

Like users and groups, IAM roles define the limits for what can be done within your AWS account. The important difference is that, unlike users and groups, roles are, for the most part, used by applications and services rather than people. Must specify exactly what permissions you want to give the role or, in other words, what you want the beneficiary processes to be able to do. From that point any authenticated mobile app users will have access to those S3 resources.

- IAM Policies
  - Conditions in IAM policies can look at resource tags to determine whether to allow a particular action. For example, you can specify a condition that permits an EC2 instance to access a production database only if the instance has the Environment tag with the value Production.

In the real world, you can attach a role to an instance that provides privileges to applications running on the instance. Roles help you avoid sharing long-term credentials like access keys and protect your instances.

- IAM Credential Reports
  - This report lists all users in your account and the status of their various credentials.
  - This lists all users in your account and the status of their various credentials, including passwords, access keys, and MFA devices. You can get a credential report from the AWS Management Console, the AWS SDKs and Command Line Tools, or the IAM API.
  - Accessed from the IAM Dashboard, a credential report displays a simple interface with no more (or less) than one lonely button: Download Report. We'll let you handle the

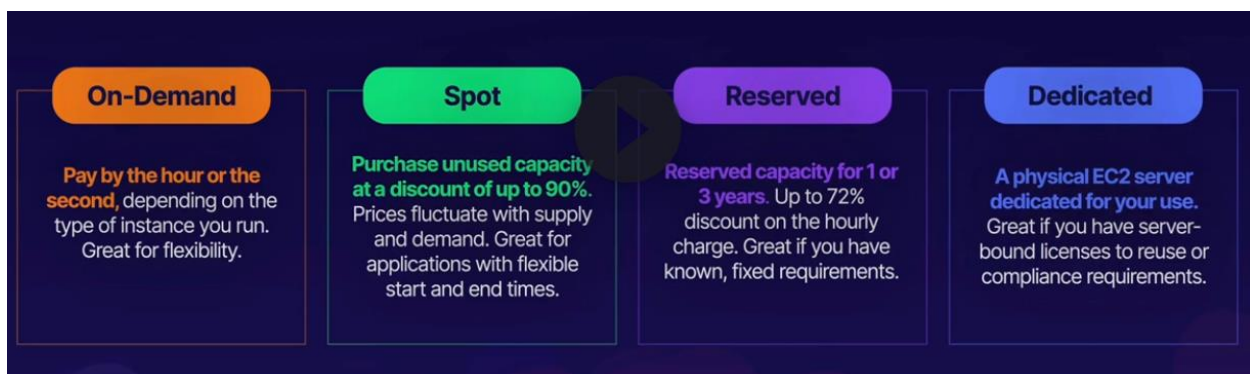
practical details from there. Accessed from the IAM Dashboard, a credential report displays a simple interface with no more (or less) than one lonely button: Download Report. We'll let you handle the practical details from there.

- Suggest downloading the comma-separated values (CSV) files the service generates.

>Security Groups and bootstrap scripts

### EC2 Exam Tips

1. EC2 is like a VM hosted in AWS instead of your own Data Center
2. Pricing Options



3. AWS Command Line
  - a. Allow only minimum account of access required to do their job
  - b. Use groups, create IAM groups and assign your users to groups

4. Secret access key
  - a. You will only see this once, so if you lose it you can delete the access key ID and secret access key and regenerate them
  - b. DO NOT SHARE KEY PAIRS
  - c. Command line supports Linux, Windows, and MACos
5. Know the placement groups
  - a. Cluster
  - b. Partition
  - c. Spread

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>