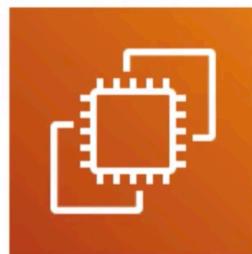


Elastic Compute Cloud (EC2)



Cloud Computing Service
Choose your **OS, Storage, Memory, Network Throughput**.
Launch and SSH into your server **within minutes**.



Introduction to EC2

Elastic Compute Cloud (EC2) is a **highly configurable server**.
EC2 is resizable **compute capacity**. It takes **minutes** to launch new instances.
Anything and everything on AWS uses EC2 Instance underneath.

Choose your OS via
Amazon Machine Image (AMI)



Choose your **Instance Type**

t2.nano

\$0.0065/hour (\$4.75/month)
1 vCPU 0.5GB Mem

C4.8xlarge

\$1.591/hour (\$1161.43/month)
36 vCPU 60GB Mem 10 Gigabit performance

Add Storage (**EBS, EFS**)

SSD **HDD** **Virtual Magnetic Tape** **Multiple Volumes**

Configure your Instance

Security Groups, Key Pairs, UserData, IAM Roles, Placement Groups



EC2 – Instance Types and Usage

General Purpose	A1 T3 T3a T2 M5 M5a M4 balance of compute, memory and networking resources Use-cases web servers and code repositories
Compute Optimized	C5 C5n C4 Ideal for compute bound applications that benefit from high performance processor Use-cases scientific modeling, dedicated gaming servers and ad server engines
Memory Optimized	R5 R5a X1e X1 High Memory z1d fast performance for workloads that process large data sets in memory. Use-cases in-memory caches, in-memory databases, real time big data analytics
Accelerated Optimized	P3 P2 G3 F1 hardware accelerators, or co-processors Use-cases Machine learning, computational finance, seismic analysis, speech recognition
Storage Optimized	I3 I3en D2 H1 high, sequential read and write access to very large data sets on local storage Use-cases NoSQL, in-memory or transactional databases, data warehousing



EC2 - Instance Sizes

EC2 Instance Sizes **generally double** in price and key attributes

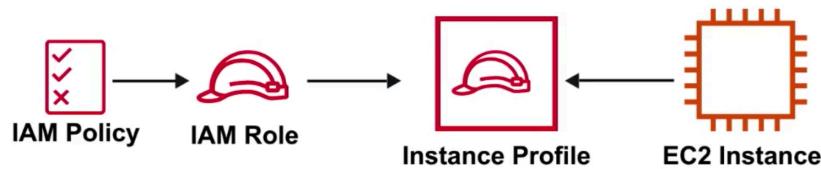
Name	vCPU	RAM (GiB)	On-Demand per hour	On-Demand per month
t2.small	1	12	\$0.023	\$16.79
t2.medium	2	24	\$0.0464	\$33.87
t2.large	2	36	\$0.0928	\$67.74
t2.xlarge	4	54	\$0.1856	\$135.48



EC2 - Instance Profile

Instead of embedding your AWS credentials (Access Key and Secret) in your code so your Instance has permissions to access certain services you can **Attach a role to an instance** via an **Instance Profile**

You want to **always avoid embedding your AWS credentials** when possible.



An **Instance Profile** holds a reference to a role. The EC2 instance is associated with the Instance Profile. When you select an IAM role when Launching an EC2 instance, AWS will automatically create the Instance Profile for you. Instance Profiles are not easily viewed via the AWS Console.

A screenshot of the AWS IAM Roles page. A red arrow points to the 'IAM role' button, which is highlighted. To its right is a small info icon. Below the button is a dropdown menu showing 'FullS3Access'. To the right of the dropdown is a 'Create new IAM role' button. At the bottom right of the screenshot is a small '(A)' icon with the word 'SUBSCRIBE' below it.



EC2 - Placement Groups

Placement Groups let you to choose **the logical placement** of your instances to optimize for **communication, performance or durability**. Placement groups are **free**.

Cluster

- packs instances close together inside an **AZ**
- low-latency network performance for tightly-coupled node-to-node communication
- well suited for High Performance Computing (HPC) applications
- Clusters cannot be multi-AZ



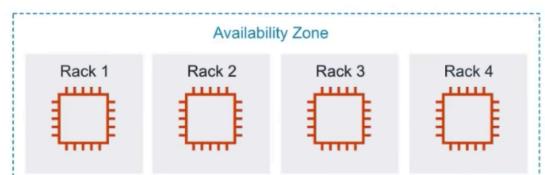
Partition

- spreads instances across logical partitions
- each partition do not share the underlying hardware with each other (rack per partition)
- well suite for large distributed and replicated workloads (Hadoop, Cassandra, Kafka)



Spread

- Each instance is placed on a different rack
- When critical instances should be keep separate from each other
- You can spread a max of 7 instances. Spreads can be multi-AZ





EC2 – UserData

You can provide an EC2 with **UserData** which is a **script** that will be automatically run when launching an EC2 instance. You could install package, apply updates or anything you like.

This example sets up an apache web-server

▼ Advanced Details

User data (i)

As text As file Input is already base64 encoded

```
#!/usr/bin/env bash
su ec2-user
sudo yum install httpd -y
sudo service httpd start
```

From within the EC2 instance, if you were to SSH in and CURL this special URL you can see the UserData script eg. [curl http://169.254.169.254/latest/user-data](http://169.254.169.254/latest/user-data)



SUBSCRIBE



EC2 - MetaData

From within your EC2 instance you can access information about the EC2 via a special url endpoint at

169.254.169.254

You would SSH into your EC2 instance and can use the CURL command:

`curl http://169.254.169.254/latest/meta-data`

/public-ipv4 get the current public IPV4 address
/ami-id the AMI ID used to launch this EC2 instance
/instance-type the Instance Type of this EC2 instance

Combine metadata with userdata scripts to perform all sorts of advanced AWS staging automation

```
[ec2-user ~]$ curl
http://169.254.169.254/latest/meta-data/
ami-id
ami-launch-index
ami-manifest-path
block-device-mapping/
events/
hostname
iam/
instance-action
instance-id
instance-type
local-hostname
local-ipv4
mac
metrics/
network/
placement/
profile
public-hostname
public-ipv4
public-keys/
reservation-id
security-groups
services/
```



EC2 CheatSheet

- **Elastic Compute Cloud (EC2)** is a Cloud Computing Service
- Configure your EC2 by choosing your **OS, Storage, Memory, Network Throughput**.
- Launch and SSH into your server **within minutes**.
- EC2 comes in variety Instance Types specialized for different roles:
 - **General Purpose** balance of compute, memory and networking resources
 - **Compute Optimized** Ideal for compute bound applications that benefit from high performance processor
 - **Memory Optimized** fast performance for workloads that process large data sets in memory.
 - **Accelerated Optimized** hardware accelerators, or co-processors
 - **Storage Optimized** high, sequential read and write access to very large data sets on local storage
- Instance Sizes **generally double** in price and key attributes
- **Placement Groups** let you to choose the logical placement of your instances to optimize for communication, performance or durability. Placement groups are free.
- **UserData** a script that will be automatically run when launching an EC2 instance.
- **MetaData** meta data about the current instance. You access this meta data via a local endpoint when SSH'd into the EC2 instance. eg. curl <http://169.254.169.254/latest/meta-data> meta data could be the instance type, current ip address etc...
- **Instance Profiles** a container for an IAM role that you can use to pass role information to an EC2 instance when the instance starts.

