

Amazon Web Services (AWS)

EC2

PRIYAN SINGH HADA

Amazon Web Services (AWS)

- ▶ Amazon Web Services (AWS) is a comprehensive cloud computing platform that includes infrastructure as a service (IaaS) and platform as a service (PaaS) offerings. AWS services offer scalable solutions for compute, storage, databases, analytics, and more.
- ▶ AWS □ Launched in 2006.
- ▶ The infrastructure consists of compute and storage servers interconnected by high-speed networks and supports a set of services.
- ▶ An application developer:
 - ▶ Installs applications on a platform of his/her choice.
 - ▶ Manages resources allocated by Amazon.

What is AWS ?

- ▶ AWS provides on-demand delivery of technology services through the Internet with pay-as-you-go pricing. This is known as cloud computing.
- ▶ The AWS Cloud encompasses a broad set of global cloud-based products that includes compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security, and enterprise applications: on-demand, available in seconds, with pay-as-you-go pricing. With over 200 fully featured services available from data centers globally, the AWS Cloud has what you need to develop, deploy, and operate your applications, all while lowering costs, becoming more agile, and innovating faster.
- ▶ For example, with the AWS Cloud, you can spin up a virtual machine, specifying the number of vCPU cores, memory, storage, and other characteristics in seconds, and pay for the infrastructure in per-second increments only while it is running. One benefit of the AWS global infrastructure network is that you can provision resources in the Region or Regions that best serve your specific use case. When you are done with the resources, you can simply delete them. With this built-in flexibility and scalability, you can build an application to serve your first customer, and then scale to serve your next 100 million.

AWS regions and availability zones

- ▶ Amazon offers cloud services through a network of data centers on several continents.
- ▶ In each *region* there are several availability zones interconnected by high-speed networks.
- ▶ An *availability zone* is a data center consisting of a large number of servers.

Region	Location	Availability zones	Cost
US West	Oregon	us-west-2a/2b/2c	Low
US West	North California	us-west-1a/1b/1c	High
US East	North Virginia	us-east-1a/2a/3a/4a	Low
Europe	Ireland	eu-west-1a/1b/1c	Medium
South America	Sao Paulo, Brazil	sa-east-1a/1b	Very high
Asia Pacific	Tokyo, Japan	ap-northeast-1a/1b	High
Asia Pacific	Singapore	ap-southeast-1a/1b	Medium

- ▶ Regions do not share resources and communicate through the Internet.

REGIONS

Services

Search

[Alt+S]

Tokyo ▲

Priyank Singh Hada ▼

Console Home [Info](#)

⋮

Welcome to AWS

⋮

[Getting started with AWS](#)

Learn the fundamentals and find valuable information to get the most out of AWS.

[Training and certification](#)

Learn from AWS experts and advance your skills and knowledge.

[What's new with AWS?](#)

Discover new AWS services, features, and Regions.

⋮

AWS Health [Info](#)

⋮

Open issues

0

Past 7 days

Scheduled changes

0

Upcoming and past 7 days

Other notifications

0

Past 7 days

[Go to AWS Health](#)

US East (N. Virginia)

us-east-1

US East (Ohio)

us-east-2

US West (N. California)

us-west-1

US West (Oregon)

us-west-2

Asia Pacific (Mumbai)

ap-south-1

Asia Pacific (Osaka)

ap-northeast-3

Asia Pacific (Seoul)

ap-northeast-2

Asia Pacific (Singapore)

ap-southeast-1

Asia Pacific (Sydney)

ap-southeast-2

Asia Pacific (Tokyo)

ap-northeast-1

Canada (Central)

ca-central-1

Europe (Frankfurt)

eu-central-1

Europe (Ireland)

eu-west-1

NOT ALLOWED REGIONS

The screenshot displays the AWS Management Console interface. The top navigation bar includes the AWS logo, a 'Services' menu, a search bar, and the user's profile 'Priyank Singh Hada' with a dropdown arrow. The main content area is titled 'Console Home' and features three primary sections: 'Welcome to AWS' with links for getting started, training, and news; 'AWS Health' showing zero open issues, scheduled changes, and other notifications; and a 'Manage Regions' dropdown menu. This menu is open, revealing a list of 10 regions that are not enabled for the account, starting with 'South America (São Paulo)'. The user's current region is 'Tokyo'.

Console Home Info

Welcome to AWS

- Getting started with AWS** [↗](#)
Learn the fundamentals and find valuable information to get the most out of AWS.
- Training and certification** [↗](#)
Learn from AWS experts and advance your skills and knowledge.
- What's new with AWS?** [↗](#)
Discover new AWS services, features, and Regions.

AWS Health Info

Open issues
0 Past 7 days

Scheduled changes
0 Upcoming and past 7 days

Other notifications
0 Past 7 days

[Go to AWS Health](#)

Manage Regions

There are 10 Regions that are not enabled for this account

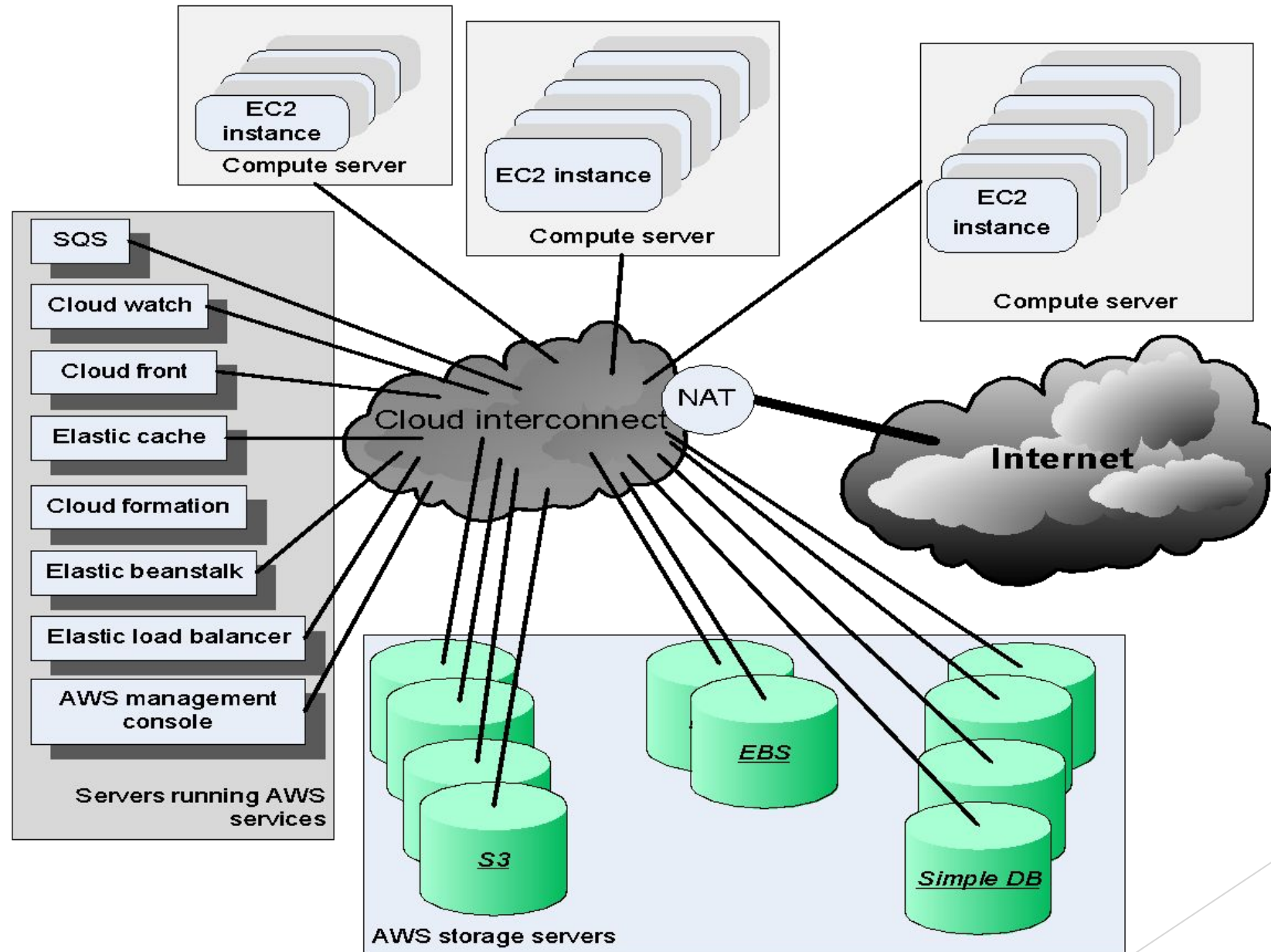
South America (São Paulo)	sa-east-1
Africa (Cape Town)	af-south-1
Asia Pacific (Melbourne)	ap-southeast-4
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Hyderabad)	ap-south-2
Asia Pacific (Jakarta)	ap-southeast-3
Europe (Milan)	eu-south-1
Europe (Spain)	eu-south-2
Europe (Zurich)	eu-central-2
Middle East (Bahrain)	me-south-1
Middle East (UAE)	me-central-1

EC2 - Elastic Cloud Computing

- ▶ *EC2* - web service for launching instances of an application under several operating systems, such as:
 - ▶ Several Linux distributions.
 - ▶ Microsoft Windows Server 2003 and 2008.
 - ▶ OpenSolaris.
 - ▶ FreeBSD.
 - ▶ NetBSD.
- ▶ A user can
 - ▶ Load an *EC2* instance with a custom application environment.
 - ▶ Manage network's access permissions.
 - ▶ Run the image using as many or as few systems as desired.

EC2 Services

- ▶ Within Amazon's Elastic Compute Cloud (EC2) offerings, clients have a choice between services:
 - ▶ On-demand instances: reserve for given price and duration
 - ▶ Spot instances: cheaper price; unpredictable duration
- ▶ Challenges:
 - ▶ Spot instance prices fluctuate a lot!
 - ▶ Duration of instance availability unknown
 - ▶ Bid failures and revocations adversely impact QoS
 - ▶ How to “optimize” cost and performance in EC2 markets?



Getting Started with Amazon EC2

- ▶ Step 1: Sign up for Amazon EC2
- ▶ Step 2: Create a key pair
- ▶ Step 3: Launch an Amazon EC2 instance
- ▶ Step 4: Connect to the instance
- ▶ Step 5: Customize the instance
- ▶ Step 6: Terminate instance and delete the volume created



Services

Search

[Alt+S]



Tokyo

Priyank Singh Hada

[Go to AWS Cost Management](#)

Build a solution [Info](#)

Start building with simple wizards and automated workflows.



[Launch a virtual machine](#)

With EC2 (2 mins)



[Start a development project](#)

With CodeStar (5 mins)



[Connect an IoT device](#)

With AWS IoT (5 mins)



[Build using virtual servers](#)

With Lightsail (2 mins)



[Host a static web app](#)

With AWS Amplify Console (2 mins)



[Register a domain](#)

With Route 53 (3 mins)



[Build a web app](#)

With AWS App Runner (5 mins)



[Deploy a serverless microservice](#)

With API Gateway (2 mins)



[Start migrating to AWS](#)

With AWS MGN (2 mins)



[Build SQL Server on AWS](#)

With high availability (HA and FCI) (2 mins)

Trusted Advisor [Info](#)

Action recommended

0

[Details](#)

Investigation recommended

0

[Details](#)

Passed items

2

[Details](#)

[Go to Trusted Advisor](#)

EC2 > Instances > Launch an instance

Launch an instance

[Info](#)

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

Name and tags

[Info](#)

Name

[Add additional tags](#)

▼ Application and OS Images (Amazon Machine Image)

[Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

▼ Summary

Number of instances

[Info](#)

Software Image (AMI)

Amazon Linux 2 Kernel 5.10 AMI...[read more](#)

ami-0bba69335379e17f8

Virtual server type (instance type)

t2.micro

Firewall (security group)

New security group

Storage (volumes)

1 volume(s) - 8 GiB

Cancel

Launch instance

Feedback

Looking for language selection? Find it in the new [Unified Settings](#)

© 2023, Amazon Web Services India Private Limited or its affiliates.

[Privacy](#)

[Terms](#)

[Cookie preferences](#)



▼ Application and OS Images (Amazon Machine Image) [Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

Recents

Quick Start

Amazon
Linux

macOS



Ubuntu



Windows



Red Hat



S

[Browse more AMIs](#)

Including AMIs from
AWS, Marketplace and
the Community

Amazon Machine Image (AMI)

Amazon Linux 2 AMI (HVM) - Kernel 5.10, SSD Volume Type
ami-0bba69335379e17f8 (64-bit (x86)) / ami-0c5cff9ad118f667a (64-bit (Arm))
Virtualization: hvm ENA enabled: true Root device type: ebs

Free tier eligible ▼

▼ Summary

Number of instances [Info](#)[Software Image \(AMI\)](#)

Amazon Linux 2 Kernel 5.10 AMI...[read more](#)
ami-0bba69335379e17f8

[Virtual server type \(instance type\)](#)

t2.micro

[Firewall \(security group\)](#)

New security group

[Storage \(volumes\)](#)

1 volume(s) - 8 GiB

Cancel

Launch instance

Amazon Elastic Compute Cloud (EC2)

- ▶ Amazon Machine Images (**AMIs**) are the basic building blocks of Amazon EC2
- ▶ An AMI is a template that contains a software configuration (operating system, application server and applications) that can run on Amazon's computing environment
- ▶ AMIs can be used to launch an ***instance***, which is a copy of the AMI running as a virtual server in the cloud.

▼ Instance type [Info](#)

Instance type

t2.micro

Free tier eligible ▼

Family: t2 1 vCPU 1 GiB Memory

On-Demand Linux pricing: 0.0152 USD per Hour

On-Demand Windows pricing: 0.0198 USD per Hour

[Compare instance types](#)▼ Key pair (login) [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name - *required*

Select ▼

[Create new key pair](#)▼ Network settings [Info](#)[Edit](#)

▼ Summary

Number of instances [Info](#)

1

[Software Image \(AMI\)](#)Amazon Linux 2 Kernel 5.10 AMI...[read more](#)

ami-0bba69335379e17f8

[Virtual server type \(instance type\)](#)

t2.micro

[Firewall \(security group\)](#)

New security group

[Storage \(volumes\)](#)

1 volume(s) - 8 GiB

Cancel

Launch instance

EC2 instances

- ▶ An instance is a virtual server with a well specified set of resources including: CPU cycles, main memory, secondary storage, communication and I/O bandwidth.
- ▶ The user chooses:
 - ▶ The region and the availability zone where this virtual server should be placed.
 - ▶ An instance type from a limited menu of instance types.
- ▶ When launched, an instance is provided with a DNS name; this name maps to a
 - ▶ *private IP address* □ for internal communication within the internal EC2 communication network.
 - ▶ *public IP address* □ for communication outside the internal Amazon network, e.g., for communication with the user that launched the instance.

AWS instances (cont'd)

- ▶ Network Address Translation (NAT) maps external IP addresses to internal ones.
- ▶ The public IP address is assigned for the lifetime of an instance.
- ▶ An instance can request an *elastic IP address*, rather than a public IP address. The elastic IP address is a static public IP address allocated to an instance from the available pool of the availability zone.
- ▶ An elastic IP address is not released when the instance is stopped or terminated and must be released when no longer needed.


AWS instances (cont'd)


- ▶ Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn more about instance types and how they can meet your computing needs.

☰

Instance types (1/506)

< 1 2 3 4 5 6 7 ... 11 >





	Instance type ▼	vCPUs ▼	Architecture ▼	Memory (GiB) ▼	Storage (GB) ▼	Storage type ▼	Network performance ▼
<input type="radio"/>	t2.nano	1	i386, x86_64	0.5	-	-	Low to Moderate
<input checked="" type="radio"/>	t2.micro	1	i386, x86_64	1	-	-	Low to Moderate
<input type="radio"/>	t2.small	1	i386, x86_64	2	-	-	Low to Moderate
<input type="radio"/>	t2.medium	2	i386, x86_64	4	-	-	Low to Moderate
<input type="radio"/>	t2.large	2	x86_64	8	-	-	Low to Moderate
<input type="radio"/>	t2.xlarge	4	x86_64	16	-	-	Moderate
<input type="radio"/>	t2.2xlarge	8	x86_64	32	-	-	Moderate
<input type="radio"/>	t3.nano	2	x86_64	0.5	-	-	Up to 5 Gigabit
<input type="radio"/>	t3.micro	2	x86_64	1	-	-	Up to 5 Gigabit



Cancel

Select instance type

☰

Instance types (1/506)

< 1 2 3 4 5 6 7 ... 11 >



	Instance type ▼	vCPUs ▼	Architecture ▼	Memory (GiB) ▼	Storage (GB) ▼	Storage type ▼	Network performance ▼
<input checked="" type="radio"/>	a1.4xlarge	16	arm64	32	-	-	Up to 10 Gigabit
<input checked="" type="radio"/>	a1.metal	16	arm64	32	-	-	Up to 10 Gigabit
<input type="radio"/>	c1.medium	2	i386, x86_64	1.7	350	hdd	Moderate
<input type="radio"/>	c1.xlarge	8	x86_64	7	1680	hdd	High
<input type="radio"/>	c3.large	2	i386, x86_64	3.75	32	ssd	Moderate
<input type="radio"/>	c3.8xlarge	32	x86_64	60	640	ssd	10 Gigabit
<input type="radio"/>	c3.xlarge	4	x86_64	7.5	80	ssd	Moderate
<input type="radio"/>	c3.4xlarge	16	x86_64	30	320	ssd	High
<input type="radio"/>	c3.2xlarge	8	x86_64	15	160	ssd	High

Cancel

Select instance type



Instance types (1/506)



Filter instance types

< 1 2 3 4 5 6 7 ... 11 > ⚙

	Instance type ▼	vCPUs ▼	Architecture ▼	Memory (GiB) ▼	Storage (GB) ▼	Storage type ▼	Network performance ▼
<input type="radio"/>	c5.metal	96	x86_64	192	-	-	25 Gigabit
<input type="radio"/>	c5a.large	2	x86_64	4	-	-	Up to 10 Gigabit
<input type="radio"/>	c5a.xlarge	4	x86_64	8	-	-	Up to 10 Gigabit
<input type="radio"/>	c5a.4xlarge	16	x86_64	32	-	-	Up to 10 Gigabit
<input type="radio"/>	c5a.12xlarge	48	x86_64	96	-	-	12 Gigabit
<input type="radio"/>	c5a.2xlarge	8	x86_64	16	-	-	Up to 10 Gigabit
<input type="radio"/>	c5a.24xlarge	96	x86_64	192	-	-	20 Gigabit
<input type="radio"/>	c5a.8xlarge	32	x86_64	64	-	-	10 Gigabit
<input type="radio"/>	c5a.16xlarge	64	x86_64	128	-	-	20 Gigabit



Cancel

Select instance type

☰

Instance types (1/506)

< 1 2 3 4 5 6 7 ... 11 >



Storage (GB) ▼	Storage type ▼	Network performance ▼	On-Demand Linux pricing ▼	On-Demand Windows pricing ▼
-	-	Very Low	0.026 USD per Hour	0.033 USD per Hour
-	-	Low to Moderate	0.0076 USD per Hour	0.0099 USD per Hour
-	-	Low to Moderate	0.0152 USD per Hour	0.0198 USD per Hour
-	-	Low to Moderate	0.0304 USD per Hour	0.0396 USD per Hour
-	-	Low to Moderate	0.0608 USD per Hour	0.0788 USD per Hour
-	-	Low to Moderate	0.1216 USD per Hour	0.1496 USD per Hour
-	-	Moderate	0.2432 USD per Hour	0.2842 USD per Hour
-	-	Moderate	0.4864 USD per Hour	0.5484 USD per Hour
-	-	Up to 5 Gigabit	0.0068 USD per Hour	0.0114 USD per Hour

Cancel

Select instance type



Subnet [Info](#)

No preference (Default subnet in any availability zone)

Auto-assign public IP [Info](#)

Enable

Firewall (security groups) [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☒ Create security group

☐ Select existing security group


We'll create a new security group called 'launch-wizard-3' with the following rules:

☒ Allow SSH traffic from
Helps you connect to your instance

Anywhere
0.0.0.0/0

☐ Allow HTTPS traffic from the internet
To set up an endpoint, for example when creating a web server

☐ Allow HTTP traffic from the internet
To set up an endpoint, for example when creating a web server

 Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.



▼ Summary

Number of instances [Info](#)

1

Software Image (AMI)

Amazon Linux 2 Kernel 5.10 AMI...[read more](#)
ami-0bba69335379e17f8

Virtual server type (instance type)

t2.micro

Firewall (security group)


New security group

Storage (volumes)

1 volume(s) - 8 GiB

Cancel

Launch instance

 Services [Alt+S]

Tokyo ▼ Priyank Singh Hada ▼

Connect to instance [Info](#)

Connect to your instance i-0cbd8f58c43e2239a using any of these options


EC2 Instance Connect

Session Manager


SSH client

EC2 serial console

Instance ID


 i-0cbd8f58c43e2239a

Public IP address

 18.179.37.175

User name

Enter the user name defined in the AMI used to launch the instance. If you didn't define a custom user name, use the default user name, ec2-user.

 **Note:** In most cases, the default user name, ec2-user, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

Cancel

Connect

[Feedback](#) Looking for language selection? Find it in the new [Unified Settings](#)

© 2023, Amazon Web Services India Private Limited or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)


```
[ec2-user@ip-172-31-43-198 ~]$
```

>

Cookie preferences

aws

Services

Search

[Alt+S]

🔍

🔔

?

Tokyo ▾

Priyank Singh Hada ▾

New EC2 Experience

Tell us what you think

EC2 Dashboard

EC2 Global View

Events

Tags

Limits

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

Images

Instances (1/3) Info

🔄

Connect

Instance state ▴

Actions ▾

Launch instances

Find instance by attribute or tag (case-sensitive)

☑

–

i-0cbd8f58c43e2239a

🟢 Running

🔍🔍

t2.m

☐

psh

i-0e5af60ccdb8da216

🟢 Running

🔍🔍

t2.m

☐

psh

i-0c51ec4d064832c76

🟢 Running

🔍🔍

t2.micro

check

Alarm status

Availability Zone

checks passed

No alarms

+

ap-northeast-1a

checks passed

No alarms

+

ap-northeast-1c

2/2 checks passed

No alarms

+

ap-northeast-1c

Stop instance

Start instance

Reboot instance

Hibernate instance

Terminate instance

Instance: i-0cbd8f58c43e2239a

⚙️

✕

Details

Security

Networking

Storage

Status checks

Monitoring

Tags

▼ Instance summary Info

Instance ID

📄 i-0cbd8f58c43e2239a

IPv6 address

–

Public IPv4 address

📄 18.179.37.175 | open address

Instance state

🟢 Running

Private IPv4 addresses

📄 172.31.43.198

Public IPv4 DNS

📄 ec2-18-179-37-175.ap-northeast-1.compute.amazonaws.com | open address

Feedback

Looking for language selection? Find it in the new Unified Settings

© 2023, Amazon Web Services India Private Limited or its affiliates.

Privacy

Terms

Cookie preferences

EC2 Instance Types

- ▶ Amazon EC2 provides a wide selection of instance types optimized to fit different use cases.
- ▶ Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications.
- ▶ Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload.

EC2 Instance TypesCont.....

- ▶ General Purpose
- ▶ Compute Optimized
- ▶ Memory Optimized
- ▶ Accelerated Computing
- ▶ Storage Optimized
- ▶ HPC Optimized

General Purpose



- ▶ General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads.
- ▶ These instances are ideal for applications that use these resources in equal proportions such as web servers and code repositories.
- ▶ Ex. M6in is suitable for network intensive requirements.
- ▶ These instances are SAP-Certified and ideal for workloads that can take advantage of high networking throughput. Workloads include high-performance file systems, distributed web scale in-memory caches, caching fleets, real-time big data analytics, Telco applications, such as 5G User Plane Function (UPF), as well as application development environments.

Instance	vCPU	Memory (GiB)	Instance Storage (GB)	Network Bandwidth (Gbps)***	EBS Bandwidth (Gbps)
m6in.large	2	8	EBS-Only	Up to 25	Up to 20
m6in.xlarge	4	16	EBS-Only	Up to 30	Up to 20
m6in.2xlarge	8	32	EBS-Only	Up to 40	Up to 20
m6in.4xlarge	16	64	EBS-Only	Up to 50	Up to 20
m6in.8xlarge	32	128	EBS-Only	50	20
m6in.12xlarge	48	192	EBS-Only	75	30
m6in.16xlarge	64	256	EBS-Only	100	40
m6in.24xlarge	96	384	EBS-Only	150	60
m6in.32xlarge	128	512	EBS-Only	200****	80

m6idn.large	2	8	1x118 NVMe SSD	Up to 25	Up to 20
m6idn.xlarge	4	16	1x237 NVMe SSD	Up to 30	Up to 20
m6idn.2xlarge	8	32	1x474 NVMe SSD	Up to 40	Up to 20
m6idn.4xlarge	16	64	1x950 NVMe SSD	Up to 50	Up to 20
m6idn.8xlarge	32	128	1x1900 NVMe SSD	50	20
m6idn.12xlarge	48	192	2x1425 NVMe SSD	75	30
m6idn.16xlarge	64	256	2x1900 NVMe SSD	100	40

Compute Optimized

C7gn	C7g	C6in	C6g	C6gn	C6i	C6a	C5	C5a	C5n	C4
------	-----	------	-----	------	-----	-----	----	-----	-----	----

- ▶ Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors.
- ▶ Instances belonging to this family are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.
- ▶ Ex. [Amazon EC2 C7gn instances](#) are powered by Arm-based AWS Graviton3E processors. They offer up to 200 Gbps of network bandwidth and up to 50% higher packet-processing performance than previous generation C6gn instances.
- ▶ Network-intensive workloads, such as network virtual appliances, data analytics, and CPU-based artificial intelligence and machine learning (AI/ML) inference

Memory Optimized

- ▶ Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.
- ▶ [Amazon EC2 R7iz instances](#) are powered by 4th Generation Intel Xeon Scalable processors and are an ideal fit for high-compute and memory-intensive workloads.
- ▶ High-compute and memory-intensive workloads such as frontend Electronic Design Automation (EDA), relational database workloads with high per-core licensing fees, and financial, actuarial, and data analytics simulation workloads.

R7iz	R6in	R6a	R6g	R6i	R5	R5a	R5b	R5n	R4	X2gd	X2idn	X2iedn
X2iezn	X1e	X1	High Memory	z1d								

Accelerated Computing

- ▶ Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs.
- ▶ [Amazon EC2 P4 instances](#) are the latest generation of GPU-based instances and provide highest performance for machine learning training and high performance computing in the cloud.
- ▶ Machine learning, high performance computing, computational fluid dynamics, computational finance, seismic analysis, speech recognition, autonomous vehicles, and drug discovery

P4	P3	P2	DL1	Trn1	Inf2	Inf1	G5	G5g	G4dn	G4ad	G3	F1	VT1
----	----	----	-----	------	------	------	----	-----	------	------	----	----	-----

Storage Optimized

Im4gn	Is4gen	I4i	I3	I3en	D2	D3	D3en	H1
-------	--------	-----	----	------	----	----	------	----

- ▶ Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.
- ▶ [Amazon EC2 Im4gn instances](#) are powered by AWS Graviton2 processors and provide the best price performance for storage-intensive workloads in Amazon EC2. They provide up to 40% better price performance, up to 44% lower cost per TB of storage over I3 instances.
- ▶ Use Cases
- ▶ These instances maximize the number of transactions processed per second (TPS) for I/O intensive and business-critical workloads which have medium size data sets and can benefit from high compute performance and high network throughput such as relational databases (MySQL, MariaDB, and PostgreSQL), and NoSQL databases (KeyDB, ScyllaDB, and Cassandra). They are also an ideal fit for workloads that require very fast access to medium size data sets on local storage such as search engines and data analytics workloads.

HPC Optimized

- ▶ High performance computing (HPC) instances are purpose built to offer the best price performance for running HPC workloads at scale on AWS.
- ▶ HPC instances are ideal for applications that benefit from high-performance processors such as large, complex simulations and deep learning workloads.
- ▶ [Amazon EC2 Hpc6id instances](#) are designed for memory-bound and data-intensive high performance computing (HPC) workloads such as finite element analysis (FEA) for crash simulations, seismic reservoir simulations, and structural simulations.



Instance Features

- ▶ Fixed Performance instance families (e.g. M6, C6, and R6) and Burstable Performance Instance families (e.g. T3)
- ▶ Burstable Performance Instances provide a baseline level of CPU performance with the ability to burst above the baseline.
- ▶ T Unlimited instances can sustain high CPU performance for as long as a workload needs it. For most general-purpose workloads, T Unlimited instances will provide ample performance without any additional charges. The hourly T instance price automatically covers all interim spikes in usage when the average CPU utilization of a T instance is at or less than the baseline over a 24-hour window. If the instance needs to run at higher CPU utilization for a prolonged period, it can do so at a flat additional charge of 5 cents per vCPU-hour.