

### **Amazon EC2**

Amazon Elastic Compute Cloud (Amazon EC2) is a web service in the AWS Compute suite of products that provides secure, resizable compute capacity in the cloud.

The EC2 simple web service interface allows us to obtain and configure capacity with minimal friction.

EC2 is designed to make web-scale cloud computing easier for developers.

Amazon EC2's simple web service interface allows us to obtain and configure capacity with minimal friction.

It provides us with complete control of our computing resources and lets us run on Amazon's proven computing environment.

Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing us to quickly scale capacity, both up and down, as our computing requirements change.

Amazon EC2 changes the economics of computing by allowing us to pay only for capacity that we actually use.

Amazon EC2 provides developers the tools to build failure resilient applications and isolate them from common failure scenarios.

### **Benefits of EC2 include:**

- **Elastic Web-Scale computing** – you can increase or decrease capacity within minutes not hours and commission one to thousands of instances simultaneously.
- **Completely controlled** – You have complete control include root access to each instance and can stop and start instances without losing data and using web service APIs.
- **Flexible Cloud Hosting Services** – you can choose from multiple instance types, operating systems, and software packages as well as instances with varying memory, CPU and storage configurations.
- **Integrated** – EC2 is integrated with most AWS services such as S3, RDS, and VPC to provide a complete, secure solution.
- **Reliable** – EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned with SLAs of 99.95% for each region.
- **Secure** – EC2 works in conjunction with VPC to provide a secure location with an IP address range you specify and offers Security Groups, Network ACLs, and IPsec VPN features.
- **Inexpensive** – Amazon passes on the financial benefits of scale by charging very low rates and on a capacity consumed basis.

**An Amazon Machine Image (AMI)** is a special type of virtual appliance that is used to create a virtual machine within the Amazon Elastic Compute Cloud (“EC2”).

An AMI includes the following:

- One or more EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance (for example, an operating system, an application server, and applications).
- Launch permissions that control which AWS accounts can use the AMI to launch instances.
- A block device mapping that specifies the volumes to attach to the instance when it's launched

AMIs come in three main categories:

- **Community AMIs** – free to use, generally you just select the operating system you want.
- **AWS Marketplace AMIs** – pay to use, generally come packaged with additional, licensed software.
- **My AMIs** – AMIs that you create yourself.

### **Metadata and User Data:**

- User data is data that is supplied by the user at instance launch in the form of a script.
- Instance metadata is data about your instance that you can use to configure or manage the running instance.
- User data is limited to 16KB.
- User data and metadata are not encrypted.
- Instance metadata is available at <http://169.254.169.254/latest/meta-data>.

The Instance Metadata Query tool allows you to query the instance metadata without having to type out the full URI or category names.

## **Pricing**

### On-demand:

- Good for users that want the low cost and flexibility of EC2 without any up-front payment or long term commitment.
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted.
- Applications being developed or tested on EC2 for the first time.

### Reserved:

- Applications with steady state or predictable usage.
- Applications that require reserved capacity.
- Users can make up-front payments to reduce their total computing costs even further.
- Standard Reserved Instances (RIs) provide up to 75% off on-demand price.
- Convertible RIs provide up to 54% off on-demand price – provides the capability to change the attributes of the RI as long as the exchange results in the creation of RIs of equal or greater value.
- Scheduled RIs are available to launch within the time window you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month.

	Standard	Convertible
Terms	1 year, 3 year	1 year, 3 year
Average discount off On-Demand price	40% - 60%	31% - 54%
Change AZ, instance size, networking type	Yes, via ModifyReservedInstance API or console	Yes, via ExchangeReservedInstance API or console
Change instance family, OS, tenancy, payment options	No	Yes
Benefit from price reductions	No	Yes

#### Spot:

- Applications that have flexible start and end times.
- Applications that are only feasible at very low compute prices.
- Users with an urgent need for a large amount of additional compute capacity.
- If Amazon terminate your instances you do not pay, if you terminate you pay for the hour.

#### Dedicated hosts:

- Physical servers dedicated just for your use.
- You then have control over which instances are deployed on that host.
- Available as On-Demand or with Dedicated Host Reservation.
- Useful if you have server-bound software licences that use metrics like per-core, per-socket, or per-VM.

- Each dedicated host can only run one EC2 instance size and type.
- Good for regulatory compliance or licensing requirements.
- Predictable performance.
- Complete isolation.
- Most expensive option.
- Billing is per host.

#### Dedicated instances:

- Virtualized instances on hardware just for you.
- Also uses physically dedicated EC2 servers.
- Does not provide the additional visibility and controls of dedicated hosts (e.g. how instance are placed on a server).
- Billing is per instance.
- May share hardware with other non-dedicated instances in the same account.
- Available as On-Demand, Reserved Instances, and Spot Instances.
- Cost additional \$2 per hour per region.

Characteristic	Dedicated Instances
Enables the use of dedicated physical servers	X
Per instance billing (subject to a \$2 per region fee)	X
Per host billing	
Visibility of sockets, cores, host ID	
Affinity between a host and instance	
Targeted instance placement	
Automatic instance placement	X
Add capacity using an allocation request	

### **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.

The table below provides an overview of the different EC2 instance types:

Category	Families	Purpose/Design
General Purpose	A1, T3, T3a, T2, M5, M5a, M4	General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads
Compute Optimized	C5, C5n, C4	Compute Optimized instances are ideal for compute bound applications that benefit from high performance processors
Memory Optimized	R5, R5a, R4, X1e, X1, High Memory, z1d	Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory
Accelerated Computing	P3, P2, G4, G3, F1	Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating-point number calculations, graphics processing, or data pattern matching
Storage Optimized	I3, I3en, D2, H1	This instance family provides Non-Volatile Memory Express (NVMe) SSD-backed instance storage optimized for low latency, very high random I/O performance, high sequential read throughput and provide high IOPS at a low cost

### **Amazon EC2 Container Service (ECS):**

Amazon Elastic Container Service (ECS) is another product in the AWS Compute category. It provides a highly scalable, high performance container management service that supports Docker containers and allows us to easily run applications on a managed cluster of Amazon EC2 instances.

Amazon ECS eliminates the need for us to install, operate, and scale our own cluster management infrastructure.



Using API calls we can launch and stop container-enabled applications, query the complete state of clusters, and access many familiar features like security groups, Elastic Load Balancing, EBS volumes and IAM roles.

Amazon ECS can be used to schedule the placement of containers across clusters based on resource needs and availability requirements.

An Amazon ECS launch type determines the type of infrastructure on which our tasks and services are hosted.

There are two launch types and the table below describes some of the differences between the two launch types:

Amazon EC2	Amazon Fargate
You explicitly provision EC2 instances	The control plane asks for resources and Fargate automatically provisions
You're responsible for upgrading, patching, care of EC2 pool	Fargate provisions compute as needed
You must handle cluster optimization	Fargate handles cluster optimization
More granular control over infrastructure	Limited control, as infrastructure is automated

**The Elastic container registry (ECR)** is a managed AWS Docker registry service for storing, managing and deploying Docker images.

There is no additional charge for Amazon ECS. You pay for AWS resources (e.g. EC2 instances or EBS volumes) you create to store and run our application.

Amazon ECR is integrated with Amazon EC2 Container Service (ECS).

With Amazon ECR, there are no upfront fees or commitments. We pay only for the amount of data we store in our repositories and data transferred to the Internet.

## **AWS Lambda**

AWS Lambda is a serverless computing technology that allows us to run code without provisioning or managing servers.

AWS Lambda executes code only when needed and scales automatically.

You pay only for the compute time you consume (you pay nothing when your code is not running).

Benefits of AWS Lambda:

- No servers to manage.
- Continuous scaling.
- Subsecond metering.
- Integrates with almost all other AWS services.

Primary use cases for AWS Lambda:

- Data processing.
- Real-time file processing.
- Real-time stream processing.
- Build serverless backends for web, mobile, IOT, and 3rd party API requests.

## **Amazon Lightsail**

Amazon Lightsail is one of the newest services in the AWS Compute suite of products. Amazon Lightsail is great for users who do not have deep AWS technical expertise as it make it very easy to provision compute services.

Amazon Lightsail provides developers compute, storage, and networking capacity and capabilities to deploy and manage websites, web applications, and databases in the cloud.

Amazon Lightsail includes everything you need to launch your project quickly – a virtual machine, SSD-based storage, data transfer, DNS management, and a static IP.

Amazon Lightsail provides preconfigured virtual private servers (instances) that include everything required to deploy and application or create a database.

The underlying infrastructure and operating system is managed by Amazon Lightsail.

Best suited to projects that require a few dozen instances or fewer.

Provides a simple management interface.

Good for blogs, websites, web applications, e-commerce etc.

Can deploy load balancers and attach block storage.

### **Public API.**

Limited to 20 Amazon Lightsail instances, 5 static IPs, 3 DNS zones, 20 TB block storage, 40 databases, and 5 load balancers per account.

Up to 20 certificates per calendar year.

Can connect to each other and other AWS resources through public Internet and private (VPC peering) networking.

Application templates include WordPress, WordPress Multisite, Drupal, Joomla!, Magento, Redmine, LAMP, Nginx (LEMP), MEAN, Node.js, and more.

Amazon Lightsail currently supports 6 Linux or Unix-like distributions: Amazon Linux, CentOS, Debian, FreeBSD, OpenSUSE, and Ubuntu, as well as 2 Windows Server versions: 2012 R2 and 2016.

### **Amazon LightSail Databases**

Amazon Lightsail databases are instances that are dedicated to running databases.

An Amazon Lightsail database can contain multiple user-created databases, and you can access it by using the same tools and applications that you use with a stand-alone database.

Amazon Lightsail managed databases provide an easy, low maintenance way to store your data in the cloud.

Amazon Lightsail manages a range of maintenance activities and security for your database and its underlying infrastructure.

Amazon Lightsail automatically backs up your database and allows point in time restore from the past 7 days using the database restore tool.

Amazon Lightsail databases support the latest major versions of MySQL. Currently, these versions are 5.6, 5.7, and 8.0 for MySQL.

Amazon Lightsail databases are available in Standard and High Availability plans.

High Availability plans add redundancy and durability to your database, by automatically creating standby database in a separate Availability Zone.

Amazon Lightsail is very affordable.

Amazon Lightsail plans are billed on an on-demand hourly rate, so we pay only for what we use.

For every Amazon Lightsail plan we use, charge us the fixed hourly price, up to the maximum monthly plan cost.