

Amazon EC2 in Cloud Computing

What is EC2?

Amazon EC2 (**Elastic Compute Cloud**) is a service provided by **AWS (Amazon Web Services)** that gives you virtual servers (called *instances*) in the cloud.

Instead of buying physical hardware (like CPUs, RAM, and storage), you can rent virtual servers on demand.

- **Elastic** → means you can scale (increase or decrease) the number of servers anytime.
- **Compute Cloud** → means it provides computing resources (CPU, RAM, storage, network) in the cloud.

Key Features of EC2

1. **On-demand computing** – pay only for the time you use.
2. **Scalability** – increase or decrease instances based on workload.
3. **Variety of instance types** – optimized for memory, compute, or storage.
4. **Security** – controlled using security groups (like a firewall).
5. **Different pricing models** – On-Demand, Reserved, Spot Instances.

Examples

Example 1: Hosting a Website

Suppose you want to launch a website. Traditionally, you would buy a server, install software, and host it.

With **EC2**:

1. Log into AWS.
2. Launch an EC2 instance (e.g., Ubuntu server).
3. Install Apache/Nginx and your website code.
4. Assign a public IP or domain.

Result: Your site is live without buying hardware.

Example 2: Data Processing

A company needs to process a large dataset once a month (e.g., financial reports).
Instead of buying a supercomputer:

- They launch 10 EC2 instances for a few hours.
- Run the data processing job.
- Terminate the instances when done.

Result: They only pay for a few hours instead of maintaining servers year-round.

Example 3: Machine Learning

You want to train a deep learning model but your laptop is too slow.
With EC2:

- Launch a GPU-optimized EC2 instance (e.g., `p3.2xlarge`).
- Train your AI model faster.
- Once done, shut down the instance.

Result: Cost-effective and powerful.

Analogy

Think of **EC2** like renting a car instead of buying one:

- If you need it for 2 hours, you rent for 2 hours.
- If you need a truck (more power), you rent a truck instead of a small car.
- When you don't need it, you return it and stop paying.

Conclusion

In short, EC2 provides flexible, scalable virtual servers in the cloud, used for websites, apps, big data, AI, and more.