

# Principles of Computer Science II

## Cloud Computing

Ioannis Chatzigiannakis

Sapienza University of Rome

Lecture 10



## AWS: Elastic Compute Cloud (EC2)

- ▶ AWS EC2 = Elastic Compute Cloud
- ▶ Resizable compute resources in the cloud.
- ▶ Minimizes the time to provision a server.
  - ▶ Introduce a new server within minimum delay.
  - ▶ Scale capacity up very fast.
- ▶ Quickly modify the capabilities of the compute instance.
  - ▶ Introduce additional computational, memory and storage capabilities.
  - ▶ Reduce computational, memory and storage capabilities.
- ▶ Shutdown - or completely remove resources.
  - ▶ Scale down very fast.
- ▶ Pay only for the resources you need.



## Typical Use Cases

- ▶ Development and Testing Environments
- ▶ Hosting of Databases
- ▶ Hosting of web services
- ▶ Data analytics
- ▶ Code repository
- ▶ GPU-assisted machine learning
- ▶ High performance computing
- ▶ Video processing
- ▶ Backup and disaster recovery
- ▶ ...



## EC2 Provisioning Options

- ▶ **On Demand** – Pay for the compute capacity by the hour.
  - ▶ No up-front payment or long-term commitment.
  - ▶ Short-term, spiky, or unpredictable workloads.
  - ▶ Applications development or testing.
- ▶ **Spot Instances** – Acquire spare capacity up to 90% off the on-demand price.
  - ▶ When start/end times are flexible.
  - ▶ Applications that are only feasible at very low compute prices.
  - ▶ Urgent computing needs for large amounts of additional capacity.
- ▶ **Reserved Instances** – Significant discount (up to 75%) compared to On-Demand instance pricing.
  - ▶ For applications that have steady state or predictable usage.
  - ▶ Long term ( $\geq 1$  year) to reduce their total computing costs.
- ▶ **Dedicated Hosts** – Physical servers dedicated for use.



## EC2 Instance Types

- ▶ **General Purpose** – balance of compute, memory and networking resources.
- ▶ **Compute Optimized** – ideal for compute bound applications that benefit from high performance processors.
- ▶ **Memory Optimized** – deliver fast performance for workloads that process large data sets in memory.
- ▶ **Accelerated computing** – 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 generic CPUs.
- ▶ **Storage optimized** – for workloads that require high, sequential read and write access to very large data sets on local storage.



## EC2 Instance Types & Resources

- ▶ **CPU** – 64-bit Arm, AMD EPYC 7000, Intel Xeon Platinum 8175M, Intel Xeon E5-2676.
  - ▶ 1 ... 192 virtual CPUs – 1 thread = 1 vCPU.
- ▶ **Memory** – 1 ... 512 GB.
- ▶ **Network** – up to 100 Gbps.
- ▶ **Storage**
  - ▶ Amazon Elastic Block Store (EBS) – easy to use, high performance block storage service.
  - ▶ 0 ... 60 TB NVMe SSD – ensure best IOPS (Input/Output operations per second).
- ▶ **Hardware Accelerators**
  - ▶ NVIDIA Tesla V100 GPUs, NVIDIA K80 GPUs, NVIDIA T4 Tensor Core GPUs.
  - ▶ AWS Inferentia Chips.
  - ▶ Xilinx Virtex UltraScale+ VU9P FPGAs



## Available OS & Software

- ▶ **Operating Systems**
  - ▶ Linux/Unix – Amazon Linux, Debian, Ubuntu, Red Hat, CentOS, SUSE, FreeBSD, Gentoo, Mint, ...
  - ▶ Windows – Server 2019, Server 2016, Server 2012.
- ▶ **Databases** – PostgreSQL, MySQL, MongoDB, Neo4J, Oracle Enterprise, Microsoft SQL, ...
- ▶ **AWS Marketplace** – a wide selection of commercial and free software from well-known vendors.



## Pricing Examples

- ▶ **General Purpose**
  - ▶ **t2.micro Linux or Windows** – 2 vCPUs + 4 GB – 750 hours free per month, \$0.05/h
  - ▶ **a1.xlarge Linux** – 4 64-bit ARM vCPUs + 8 GB – \$0.1152/h
  - ▶ **a1.xlarge Linux** – 4 64-bit ARM vCPUs + 8 GB – \$0.1152/h
  - ▶ **m5.24xlarge Linux** – 96 Xeon vCPUs + 337 GB – \$5.136/h
  - ▶ **m5.24xlarge Windows** – 96 Xeon vCPUs + 337 GB – \$9.552/h
- ▶ **Compute Optimized**
  - ▶ **c5.xlarge Linux** – 4 Xeon vCPUs + 8 GB – \$0.192/Hour
  - ▶ **c5.24xlarge Linux** – 96 Xeon vCPUs + 192 GB – \$4.608/Hour
- ▶ **Hardware Accelerators**
  - ▶ **p3.2xlarge Linux** – 1 NVIDIA Tesla V100 GPUs + 8 Xeon vCPUs + 61 GB – \$3.305 per Hour
  - ▶ **p3dn.24xlarge Linux** – 8 NVIDIA Tesla V100 GPUs + 96 Xeon vCPUs + 768 GB – \$33.711 per Hour



## Amazon Elastic Block Store (EBS)

- ▶ Easy to use, high performance block storage service.
- ▶ Targeting both throughput and transaction intensive workloads.
  - ▶ Can be used for relational and non-relational databases.
  - ▶ Enterprise applications.
  - ▶ Big data analytics engines.
  - ▶ General purpose file systems.
  - ▶ Media workflows.
- ▶ Highly availability and durability – 99.999%
- ▶ Virtually unlimited scale – as little as a single GB of storage, or scale up to petabytes of data.
- ▶ Secure – encryption of data at-rest, data in-transit, and all volume backups.



## EBS Volume Types – HDD based

- ▶ **Throughput Optimized HDD (ST1)** – ideal for frequently accessed, throughput-intensive workloads.
  - ▶ Large datasets and large I/O sizes, such as MapReduce, Kafka, log processing, data warehouse, and ETL workloads.
  - ▶ Low cost HDD volume.
  - ▶ Volume Size: 500 GB – 16 TB.
  - ▶ Max IOPS/Volume: 500
  - ▶ Max Throughput/Volume: 500 MB/s
  - ▶ Price: \$0.045/GB-month
- ▶ **Low-cost HDD (SC1)** – ideal for less frequently accessed workloads with large, cold datasets.
  - ▶ Colder data requiring fewer scans per day.
  - ▶ Volume Size: 500 GB – 16 TB.
  - ▶ Max IOPS/Volume: 250
  - ▶ Max Throughput/Volume: 250 MB/s
  - ▶ Price: \$0.025/GB-month



## EBS Volume Types – SSD based

- ▶ **Provisioned IOPS SSD (IO1)** – high performance SSD volume designed for latency-sensitive transactional workloads.
  - ▶ I/O-intensive NoSQL & relational databases.
  - ▶ Volume Size: 4 GB – 16 TB.
  - ▶ Max IOPS/Volume: 64,000
  - ▶ Max Throughput/Volume: 1,000 MB/s
  - ▶ Price: \$0.125/GB-month + \$0.065/provisioned IOPS
- ▶ **Default EBS volume type (GP2)** – ideal for suitable for a broad range of transactional workloads.
  - ▶ Boot volumes, low-latency interactive apps, dev & test.
  - ▶ Volume Size: 1 TB – 16 TB.
  - ▶ Max IOPS/Volume: 16,000
  - ▶ Max Throughput/Volume: 250 MB/s
  - ▶ Price: \$0.10/GB-month



## Choose Region

The screenshot shows the AWS Management Console interface. On the right side, there is a 'Choose Region' dropdown menu. The 'us-east-1' region is highlighted with a red circle. Other visible regions include us-east-2, us-west-1, us-west-2, sa-east-1, eu-north-1, eu-west-1, eu-west-2, eu-west-3, eu-central-1, eu-central-2, eu-south-1, eu-south-2, eu-south-3, eu-south-4, eu-south-5, eu-south-6, eu-south-7, eu-south-8, eu-south-9, eu-south-10, eu-south-11, eu-south-12, eu-south-13, eu-south-14, eu-south-15, eu-south-16, eu-south-17, eu-south-18, eu-south-19, eu-south-20, eu-south-21, eu-south-22, eu-south-23, eu-south-24, eu-south-25, eu-south-26, eu-south-27, eu-south-28, eu-south-29, eu-south-30, eu-south-31, eu-south-32, eu-south-33, eu-south-34, eu-south-35, eu-south-36, eu-south-37, eu-south-38, eu-south-39, eu-south-40, eu-south-41, eu-south-42, eu-south-43, eu-south-44, eu-south-45, eu-south-46, eu-south-47, eu-south-48, eu-south-49, eu-south-50, eu-south-51, eu-south-52, eu-south-53, eu-south-54, eu-south-55, eu-south-56, eu-south-57, eu-south-58, eu-south-59, eu-south-60, eu-south-61, eu-south-62, eu-south-63, eu-south-64, eu-south-65, eu-south-66, eu-south-67, eu-south-68, eu-south-69, eu-south-70, eu-south-71, eu-south-72, eu-south-73, eu-south-74, eu-south-75, eu-south-76, eu-south-77, eu-south-78, eu-south-79, eu-south-80, eu-south-81, eu-south-82, eu-south-83, eu-south-84, eu-south-85, eu-south-86, eu-south-87, eu-south-88, eu-south-89, eu-south-90, eu-south-91, eu-south-92, eu-south-93, eu-south-94, eu-south-95, eu-south-96, eu-south-97, eu-south-98, eu-south-99, eu-south-100.



# Open EC2 Service Dashboard

The screenshot shows the AWS Management Console interface. In the left-hand navigation pane, the 'EC2' service is highlighted with a red circle. The main content area displays the 'AWS services' section with a search bar and a list of services. The 'EC2' service is listed as 'Amazon Elastic Compute Cloud'.

# Launch Instance

The screenshot shows the 'Launch Instance' wizard in the AWS Management Console. The 'Launch Instance' section is highlighted with a red circle, and the 'Launch Instance' button is also highlighted with a red circle. The wizard is in the 'Select an Amazon Machine Image (AMI)' step.

# Select Amazon Machine Image (AMI)

The screenshot shows the 'Select an Amazon Machine Image (AMI)' page in the AWS Management Console. The 'Amazon Linux 2 AMI' is selected, and the 'Select' button is highlighted with a red circle. The page displays a list of AMIs with their respective details and 'Select' buttons.

# Choose Instance Type

The screenshot shows the 'Choose Instance Type' page in the AWS Management Console. The 'View Configurations and Details' button is highlighted with a red circle. The page displays a table of instance types with their respective details and 'View Configurations and Details' buttons.

Family	Instance	vCPUs	Memory (GB)	Network (Mbps)	EMI supported	Network Performance	High-Speed Network
General purpose	T3	1	8.0	100	Yes	Low to Moderate	Yes
General purpose	T3a	1	8.0	100	Yes	Low to Moderate	Yes
General purpose	T3a.xlarge	4	32.0	100	Yes	Low to Moderate	Yes
General purpose	T3.xlarge	4	32.0	100	Yes	Low to Moderate	Yes
General purpose	T3.medium	2	4.0	100	Yes	Low to Moderate	Yes
General purpose	T3.large	2	8.0	100	Yes	Low to Moderate	Yes
General purpose	T3.xl	8	32.0	100	Yes	High	Yes
General purpose	T3a.xl	8	32.0	100	Yes	High	Yes
General purpose	T3a.2xlarge	8	64.0	100	Yes	High	Yes
General purpose	T3.2xlarge	8	64.0	100	Yes	High	Yes
General purpose	T3a.4xlarge	16	128.0	100	Yes	High	Yes
General purpose	T3.4xlarge	16	128.0	100	Yes	High	Yes
General purpose	T3a.8xlarge	32	256.0	100	Yes	High	Yes
General purpose	T3.8xlarge	32	256.0	100	Yes	High	Yes
General purpose	T3a.16xlarge	64	512.0	100	Yes	High	Yes
General purpose	T3.16xlarge	64	512.0	100	Yes	High	Yes

## Configure Instance Details

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the console, but, request type instances take advantage of Amazon's pricing, so you can access management tools like the console, and more.

Number of instances: 1

Placement group:  Request spot instances

Network:  Request IPv6

Subnet: us-east-1-subnet-1

Elastic IP: None

Capacity reservation:  Request capacity reservation

Subnet: us-east-1-subnet-1

IPAM:  Request IPAM

IPAM: us-east-1

Subnet: us-east-1-subnet-1

Step: Network interface

Enable secondary private IP:  Enable secondary private IP

Workgroup:  Enable Amazon EMRFS workgroup

Network interface: Elastic Network Adapter

Elastic IP: None

Advanced Details

Network interface: Elastic Network Adapter

Elastic IP: None

Next: Add Storage

## Configure Instance Details

Step 3: Configure Instance Details

Network interface: Elastic Network Adapter

Enable secondary private IP:  Enable secondary private IP

Workgroup:  Enable Amazon EMRFS workgroup

Network interface: Elastic Network Adapter

Elastic IP: None

Advanced Details

Network interface: Elastic Network Adapter

Elastic IP: None

Next: Add Storage

## Add Storage

Step 4: Add Storage

Your instance will be launched with the following storage device settings. These values are additional EBS volumes and instance store volumes to your instance, in addition to the root volume. You can also attach additional EBS volumes after launching an instance for additional storage volumes. Consideration about storage classes in Amazon EC2.

Volume Type	Size (GB)	Requester	File System	Volume Type	Size (GB)	Throughput (MB/s)	Index on Provisioning	Encryption
Root	16	amazon	ext4	Standard EBS (gp2)	100	300	None	Not Encrypted

Add New Storage

How do eligible instances can get up to 100 TB of EBS Standard Provisioned IOPS or highest storage class from Amazon SageMaker Elastic and SageMaker Inference?

Next: Add Tags

## Add Tags

Step 5: Add Tags

Any number of user-defined tags can be added. For example, you could tag by key: Name and value: MyInstance. A list of the tags you can add is shown below. Tags will be applied to all instances and volumes EBS resources.

Key	Value
aws:cloudwatch:log-group-name	aws-logs-2014-11-13-us-east-1

Add New Tag

Next: Configure Security Groups

## Configure Security Group

This screenshot shows the 'Configure Security Group' step in the AWS IAM console. The page title is 'Step 4: Configure Security Group'. It includes a warning box at the top stating: 'Warning: Make sure you specify an IP or IPv6 address to restrict your instance. We recommend setting security group rules to allow access from IPv4 or IPv6 only.' Below this, there is a form to create a security group. The 'Security group name' is 'my-security-group' and the 'Description' is 'my-security-group'. A table below shows the configuration: Type: 'sg', Protocol: 'TCP', Port Range: '20', Owner: 'self', and Description: 'my-security-group'. At the bottom right, the 'Review and Launch' button is circled in red.

## Launch Instance

This screenshot shows the 'Review Instance Launch' step in the AWS IAM console. The page title is 'Step 7: Review Instance Launch'. It includes a warning box at the top stating: 'Warning: Improve your instance security. Your security group, launch-template, is open to the world. Your relationship to the instance from IP address, the relationship to its ability to reach your security rules is also open from IPv4 or IPv6 address only. You can also open additional policies your security group to facilitate access to the application or service you're testing, e.g., HTTP, SMTP for web access. IAM security group'. Below this, there are sections for 'IAM Details', 'Instance Type', 'Security Groups', and 'Instance Details'. The 'Review and Launch' button at the bottom right is circled in red.

## Create Key pair

This screenshot shows the 'Review Instance Launch' step in the AWS IAM console. A dialog box titled 'Select an existing key pair or create a new key pair' is open. The dialog contains the text: 'You can choose an existing key pair that exists and is attached to the instance profile. When you are connected to your instance console, the instance will be prompted to allow the connection and then you will be able to connect to the instance. When you are not connected to your instance, you will be prompted to allow the connection to the instance. When you are not connected to your instance, you will be prompted to allow the connection to the instance. Note: The relationship you will be able to access the set of resources authorized by this instance when you are working on the instance console only.' Below this text, there is a dropdown menu for 'Create a new key pair' with 'my-key-pair' selected. The 'Create' button is circled in red. At the bottom right of the main page, the 'Review and Launch' button is also circled in red.

## Instance Ready to Launch

This screenshot shows the 'Launch Status' page in the AWS IAM console. The page title is 'Launch Status'. It features a large blue circular icon with a refresh symbol and the text 'Initiating Instance Launches'. Below this, there is a list of instance launches with columns for 'Instance ID', 'Instance Name', 'Instance Type', 'Instance Profile', 'Instance State', and 'Instance Details'. The page is currently empty of data.

## List of Instances

The screenshot displays the 'Instances' page in the AWS Management Console. A table lists several EC2 instances. The instance 'i-078657210625649' is highlighted in blue. The table columns include Instance ID, Name, Availability Zone, Instance Type, Status Check, Action Links, Public IPv4 DNS, and Public IPv6 DNS.

Instance ID	Name	Availability Zone	Instance Type	Status Check	Action Links	Public IPv4 DNS	Public IPv6 DNS
i-078657210625649	Ubuntu-EC2-Insta	us-east-1a	t3.micro	OK	Stop Instance	ec2-54-93-24-104.us-east-1a.elb.amazonaws.com	fd00:10:10:104::104

## Overview of Instance

The screenshot shows the 'Overview' page for the instance 'i-078657210625649'. It provides detailed information about the instance's configuration, including its name, type, status, and network settings. A 'Connect' button is visible at the top right.

**Instance summary:** Name: i-078657210625649, Type: t3.micro, Status: Running, Public IPv4 DNS: ec2-54-93-24-104.us-east-1a.elb.amazonaws.com, Public IPv6 DNS: fd00:10:10:104::104.

## Connect to Instance

The screenshot shows the 'Connect to Instance' dialog box. It contains fields for 'Instance ID' (i-078657210625649) and 'SSH key pair' (my-key-pair). A red circle highlights the 'Connect' button at the bottom right of the dialog.

## Command Line Console

The screenshot shows the 'Command Line Console' for the instance. It displays the output of the 'cat /etc/passwd' command, showing the system's user database. The output includes the root user and several system users like bin, daemon, and nobody.

```
cat /etc/passwd
root:x:0:0:root:/root:/bin/bash
bin:x:1:1:bin:/bin:/usr/sbin/nologin
daemon:x:2:2:daemon:/sbin:/usr/sbin/nologin
...
nobody:x:65534:65534:nobody:/nonexistent:/usr/sbin/nologin
```

## Other Examples

- ▶ Start, Stop, Terminate instance.
- ▶ Change Instance Type.
- ▶ Add Storage Volumes.
- ▶ Configure Security Groups.

## Connecting to the Instance

