

NAYANA DAVIS

---

# AWS: AMAZON WEB SERVICES

## LEARNING OBJECTIVES

- ▶ Explain what services AWS offers and which ones are relevant to data science
- ▶ Start and terminate an EC2 instance in the cloud
- ▶ Store and download data from an S3 bucket

## INTRO

# What are Amazon Web Services (AWS)?

## INTRO

- ▶ Cloud computing services
- ▶ Essentially, virtual machines somewhere in a datacenter that you can access and pay only for the time you need them
- ▶ AWS now has more than 70 services that span a wide range including compute, storage, networking, database, analytics, application services, deployment, management, mobile, developer tools and tools for the Internet of things.
- ▶ Main ones for data science: Elastic Compute Cloud (EC2) & Simple Storage Service (S3)

## KEY TERMINOLOGY

- ▶ Instance: virtual machine hosted in Amazon Cloud running the software we want
- ▶ Amazon Machine Image (AMI): An Amazon Machine Image (AMI) provides the information required to launch an instance, which is a virtual server in the cloud. You specify an AMI when you launch an instance, and you can launch as many instances from the AMI as you need. You can also launch instances from as many different AMIs as you need.
- ▶ SSH Key: pair of keys necessary to connect to an instance remotely. The private key will be downloaded to our laptop, the matching public key will be automatically configured on the instance.

## ADVANTAGES OF USING A SERVER IN THE CLOUD INSTEAD OF MANAGING OUR OWN DATA CENTER

- ▶ Cost reduction: don't pay infrastructure costs when you don't need it
- ▶ Reliability: Servers are maintained and guaranteed by a company whose only job is to make sure they are available for you
- ▶ Scalability: Can add more computing power when necessary

TEXT

---

**CREATE AN AWS ACCOUNT**

5 minutes

## EC2

- ▶ EC2 is a pay-as-you-go cloud platform that includes computing power, storage and database services as its core components.
- ▶ Users select the CPU(central processing unit), memory, storage, networking capacity and access controls, OS, security, and additional software needed to run a virtual computing environment, or instance.



TEXT

---

## EC2 TUTORIAL

15 minutes

## ADDITIONAL POINTS

- ▶ Pricing: EC2 pricing depends on the type of instance and on the chosen region.
- ▶ Spot Instances: Spot instances enable you to bid on unused EC2 instances, which can lower your Amazon EC2 costs significantly. The hourly price for a Spot instance (of each instance type in each Availability Zone) is set by Amazon EC2, and fluctuates depending on the supply of and demand for Spot instances. Your Spot instance runs whenever your bid exceeds the current market price.
- ▶ More info on choosing regions: <http://cloudacademy.com/blog/how-to-choose-the-best-aws-region-for-your-workload/>

## S3

- ▶ A way to store data in the cloud
- ▶ It provides storage through web services interfaces (REST, SOAP, and BitTorrent) using an object storage architecture
- ▶ Objects are organized into buckets (each owned by an Amazon Web Services account), and identified within each bucket by a unique, user-assigned key. Buckets and objects can be created, listed, and retrieved using either a REST-style HTTP interface or a SOAP interface.
- ▶ Think of it like Dropbox

TEXT

---

## S3 TUTORIAL

15 minutes

# AWSCLI

- ▶ Amazon Webservices Command Line Interface
- ▶ AWSCLI is a unified command line interface to Amazon Web Services. It allows us to control most of AWS services from the same command line interface.

TEXT

---

# AWSCLI TUTORIAL

15 Minutes

## EC2 FROM THE COMMAND LINE

15 Minutes