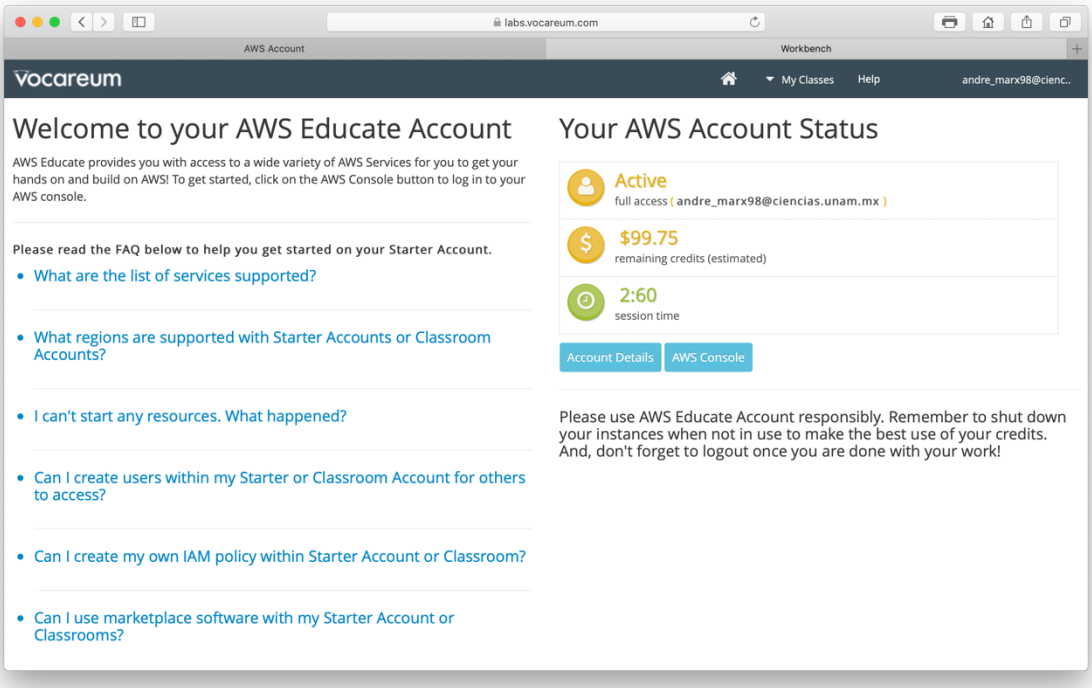
Universidad Nacional Autónoma de México

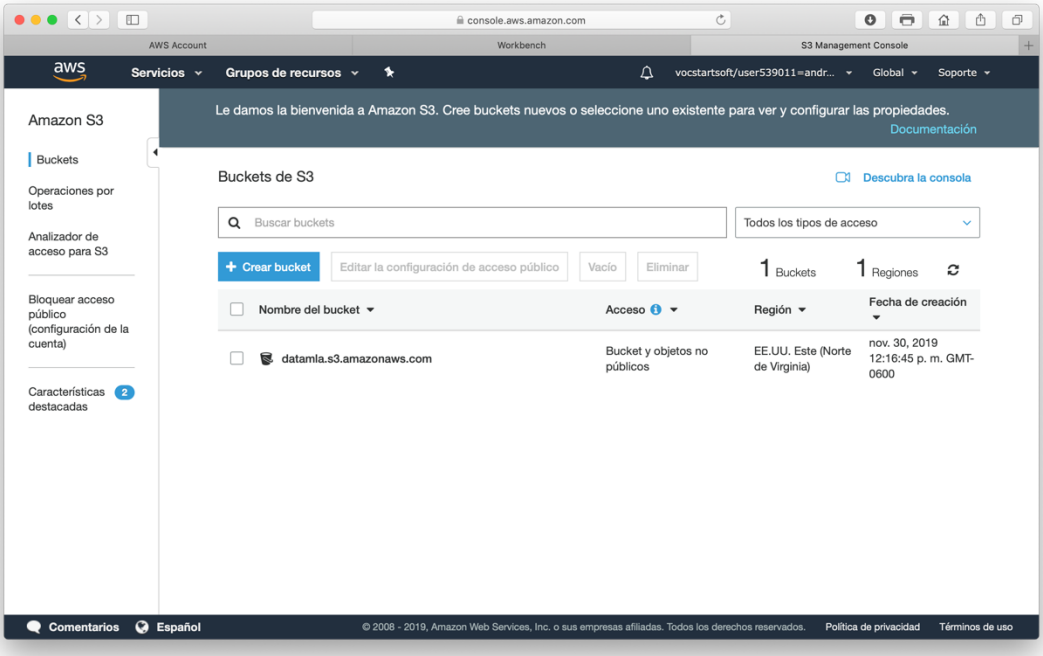
Práctica de Cloud

Prof. Jimmy Hernández Morales
Ayudante: Miguel Hinojosa Medrano
Alumno: André Marx Puente Arévalo

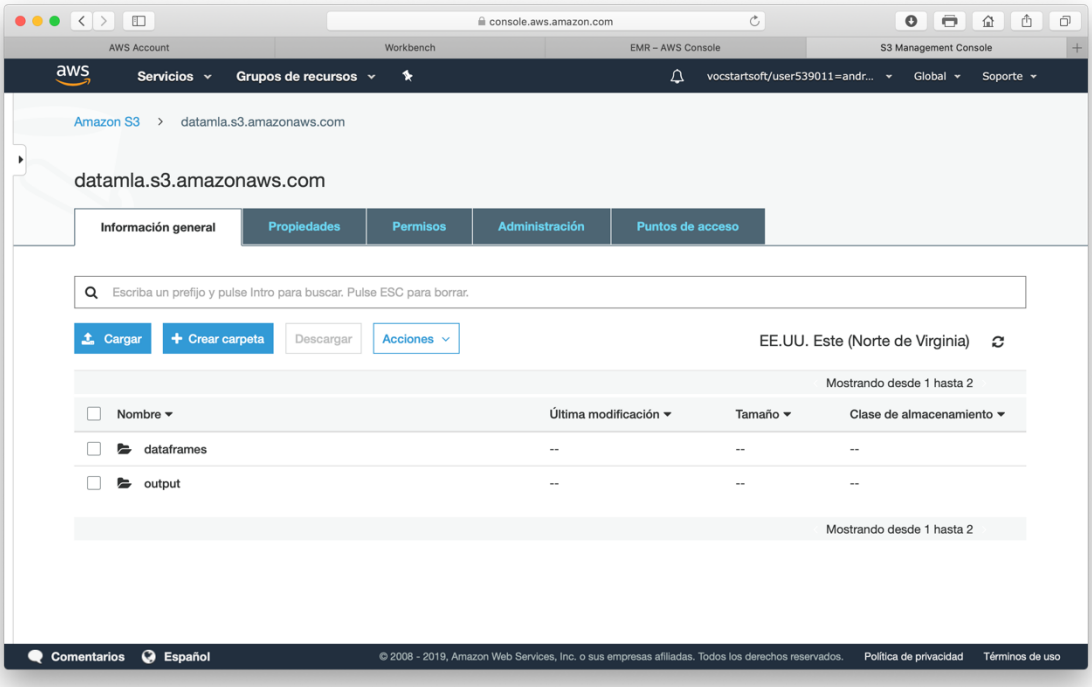
1. Abrimos la consola de AWS



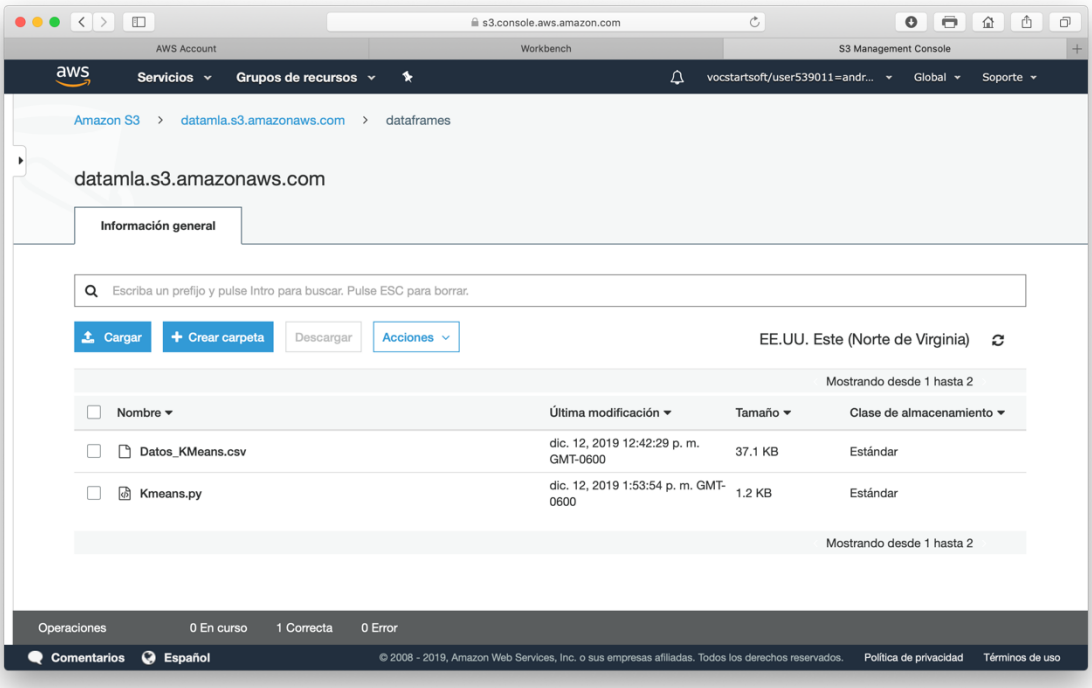
2. Creamos un bucket con el siguiente formato: “datamla.s3.amazonaws.com”



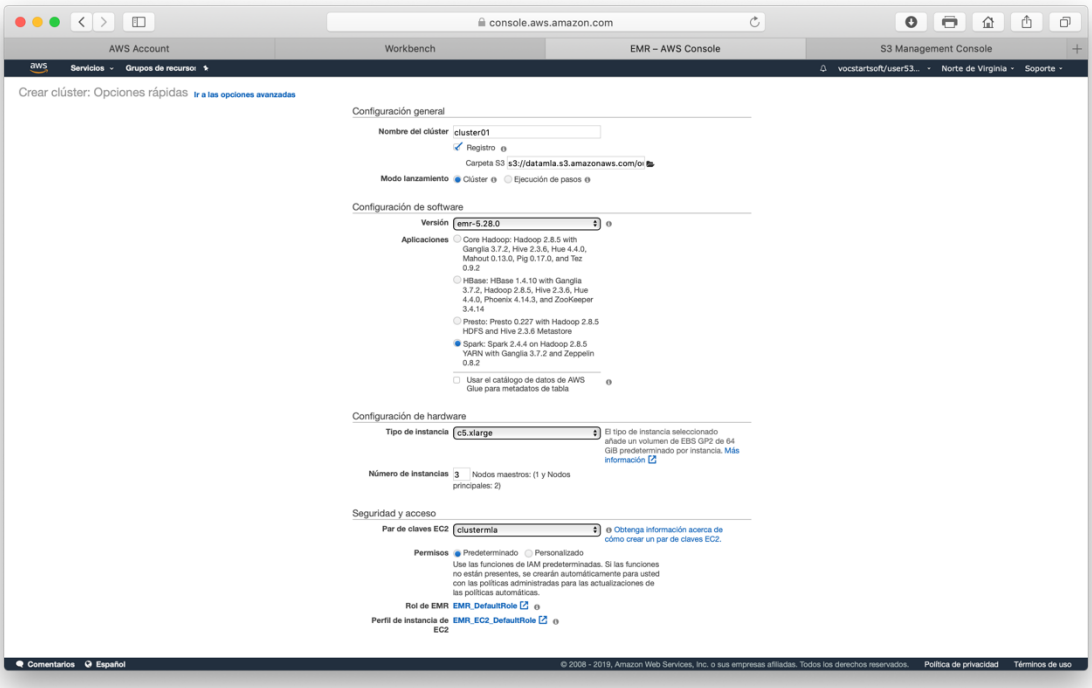
3. Creamos dos carpetas dentro del bucket “dataframes” que contendrán el archivo csv y el spcrip y “output” donde se depositarán los logs del procesamiento.

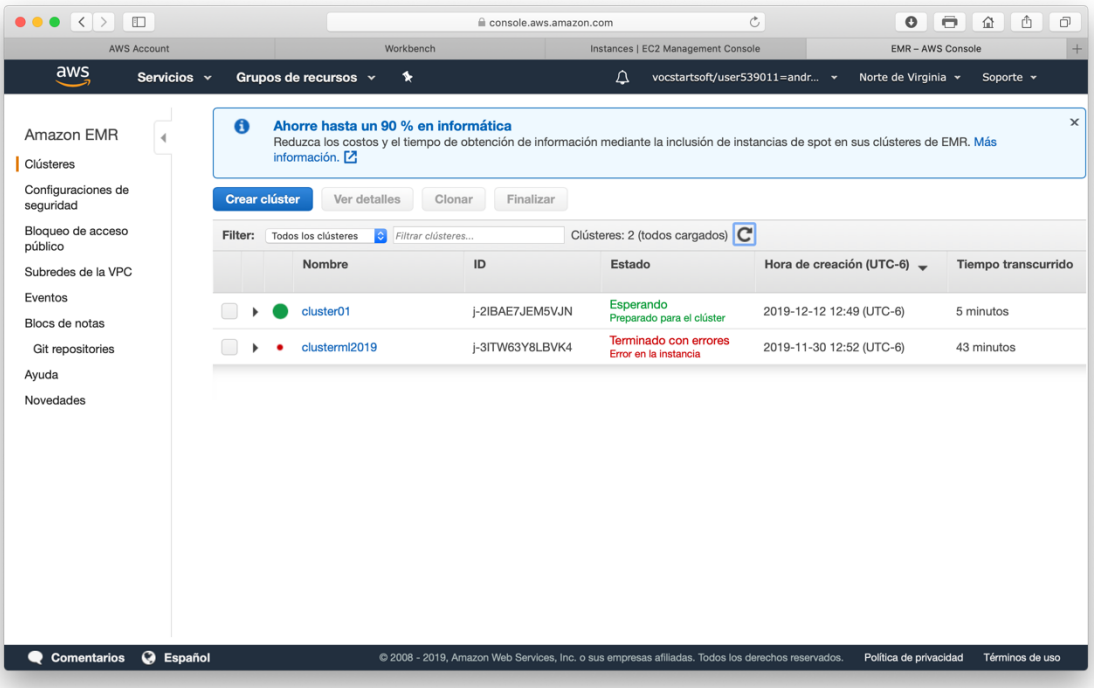


4. Almacenamos los archivos que vamos a utilizar: “Datos_KMeans.csv” y “Kmeans.py” en S3

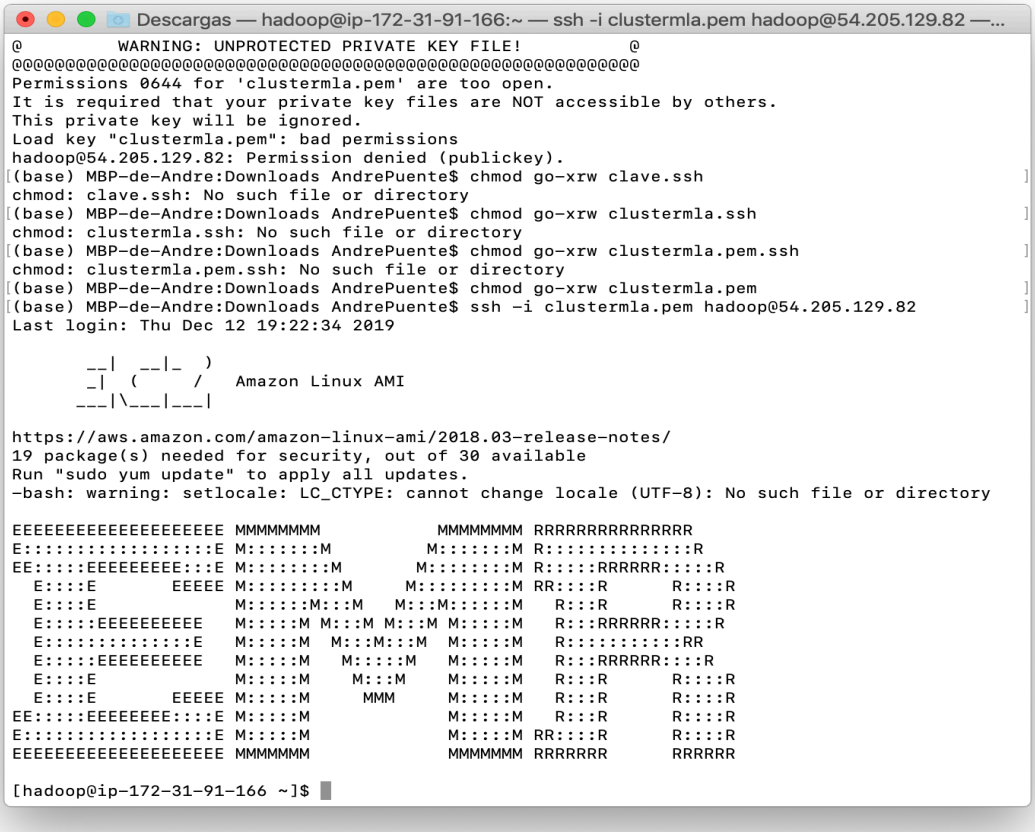


5. Lanzo un clúster con Elastic Map Reduce EMR, el cual contiene 1 master y 2 esclavos

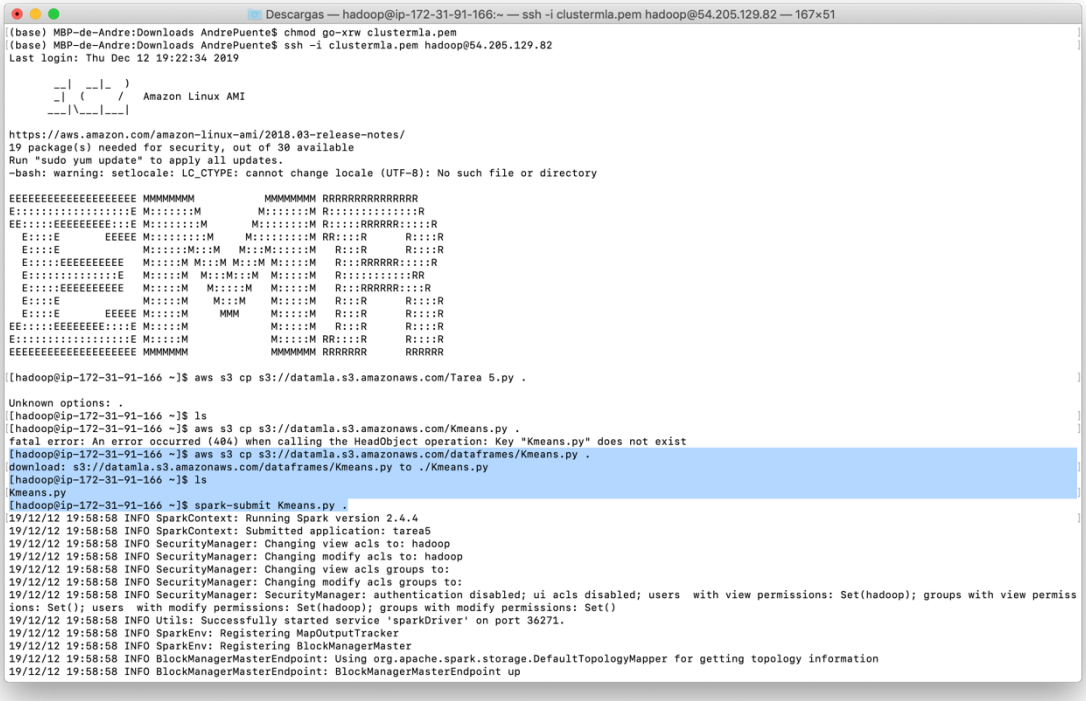


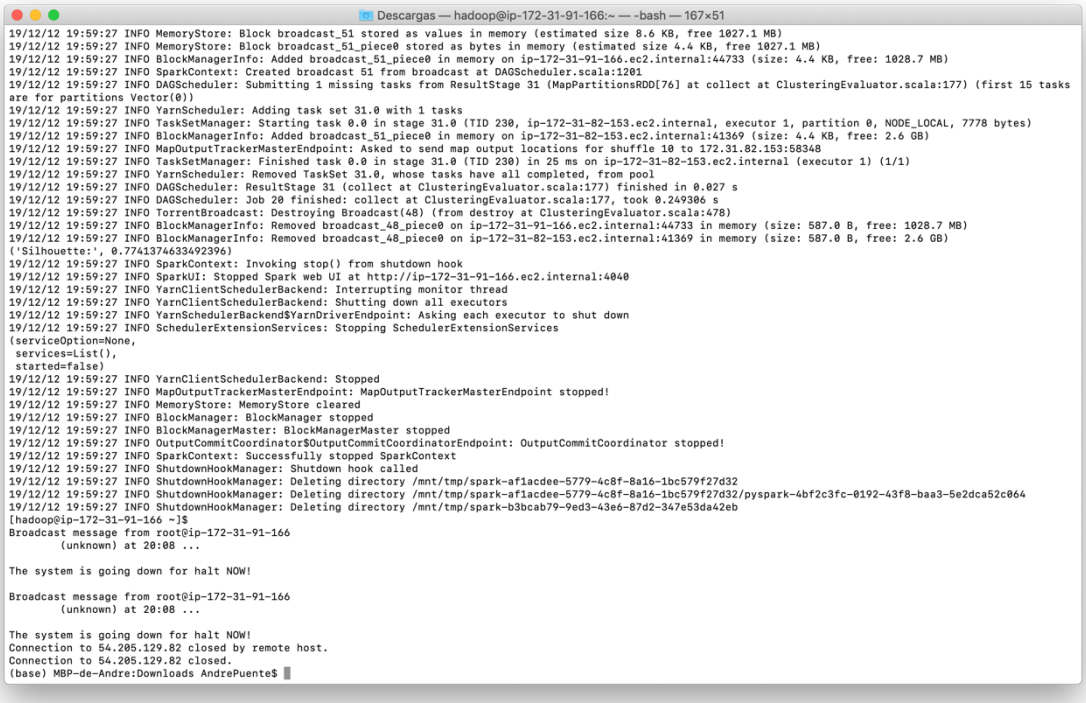


6. Me conecto al master



7. Copio mi archivo “Kmeans.py” a la instancia EC2 master y finalmente lanzo el spark submit





Nota: Lo último que sale en la terminal (imagen de arriba) es que detuve las instancias para que no me siguieran gastando créditos, pero si notamos, corrió todo el código, el cual terminaba cuando pintaba el Silhouette y este sale en la imagen.