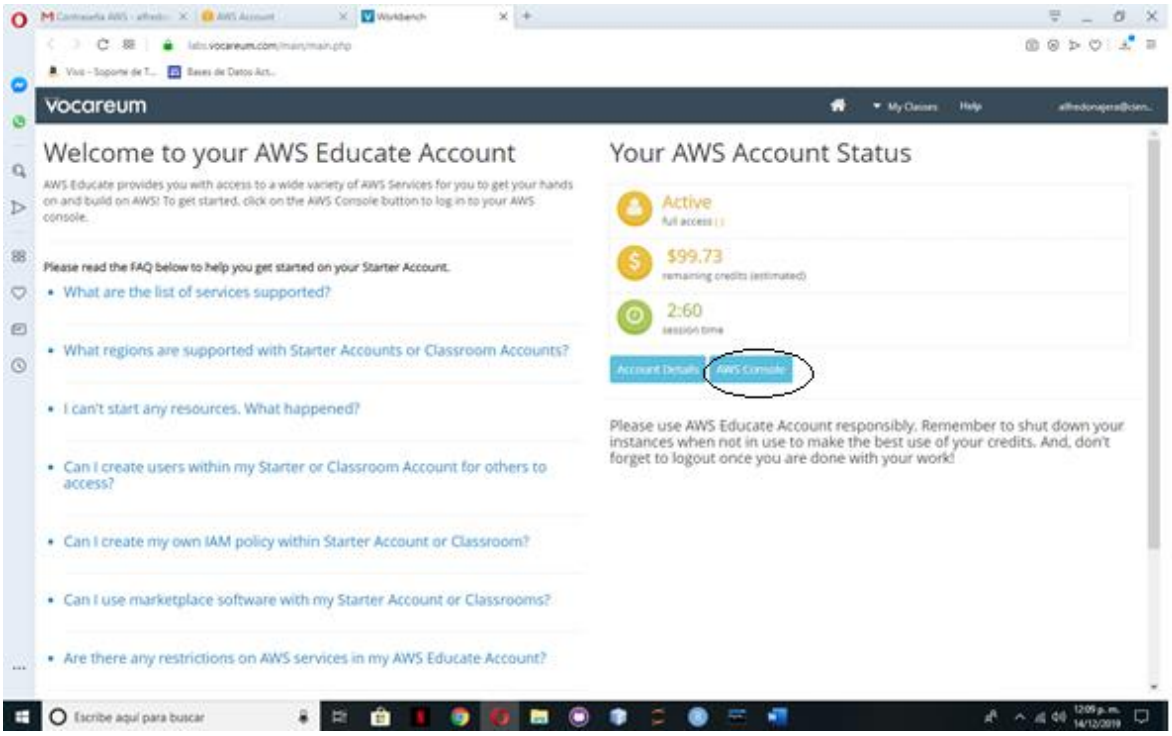
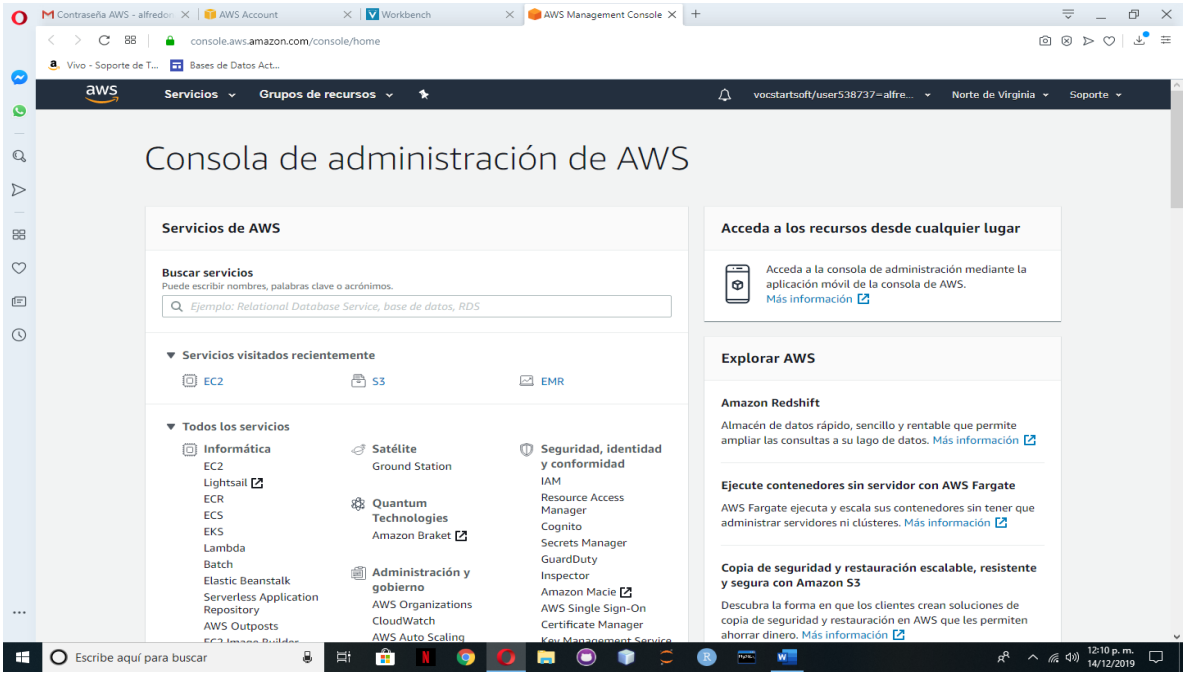
Nájera Nájera Alfredo 12 Diciembre 2019

Tarea AWS

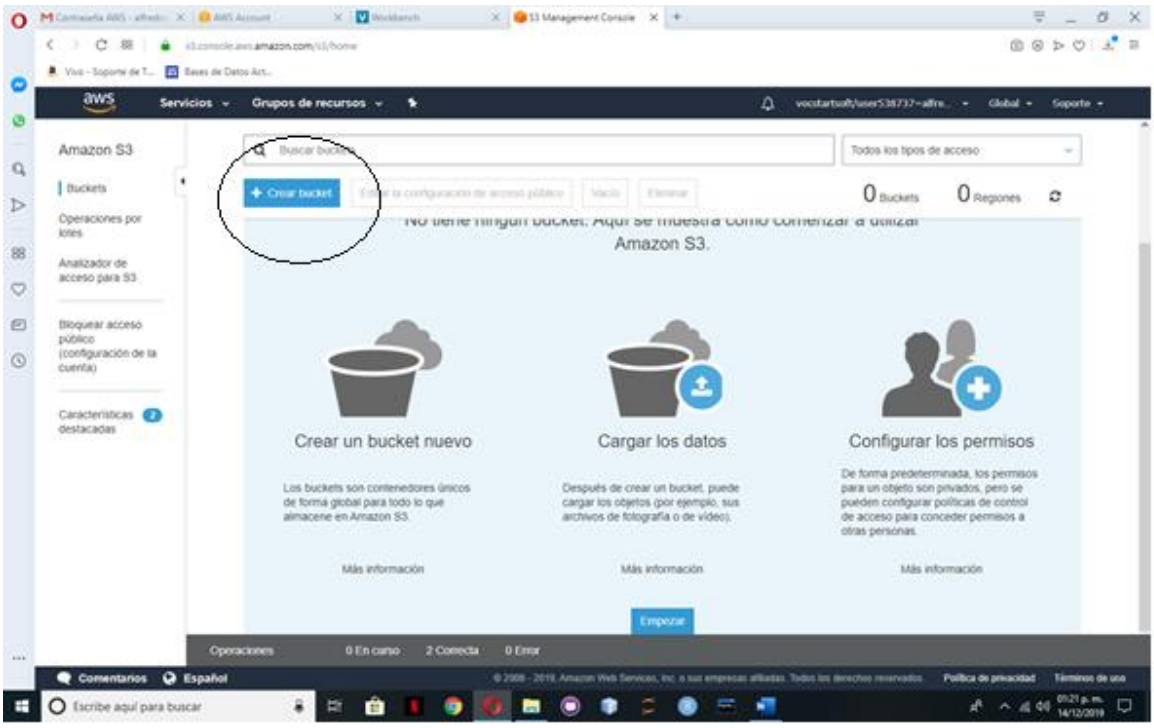
Pregunta 1

1.- Abrimos la consola de AWS después de ingresar a nuestra cuenta para poder buscar lo que es S3 (el bucket), EC2(Instancia) y EMR (Para usar spark).

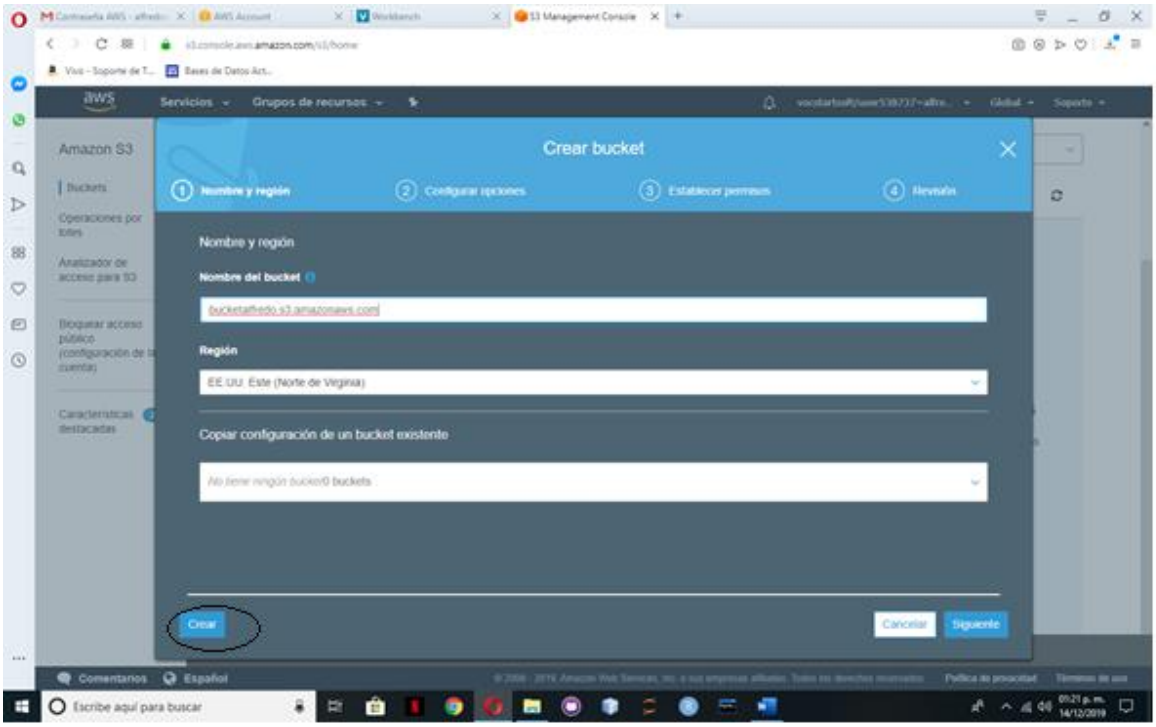




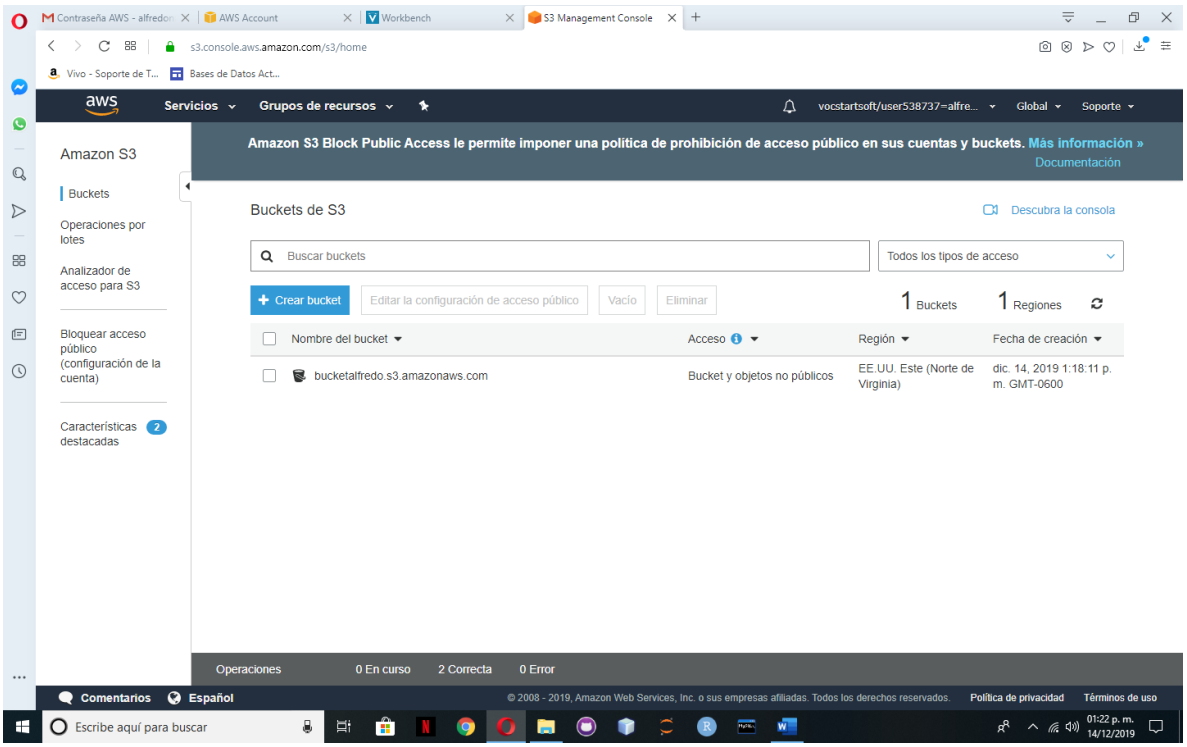
2.- Creamos un bucket buscandolo en la consola.



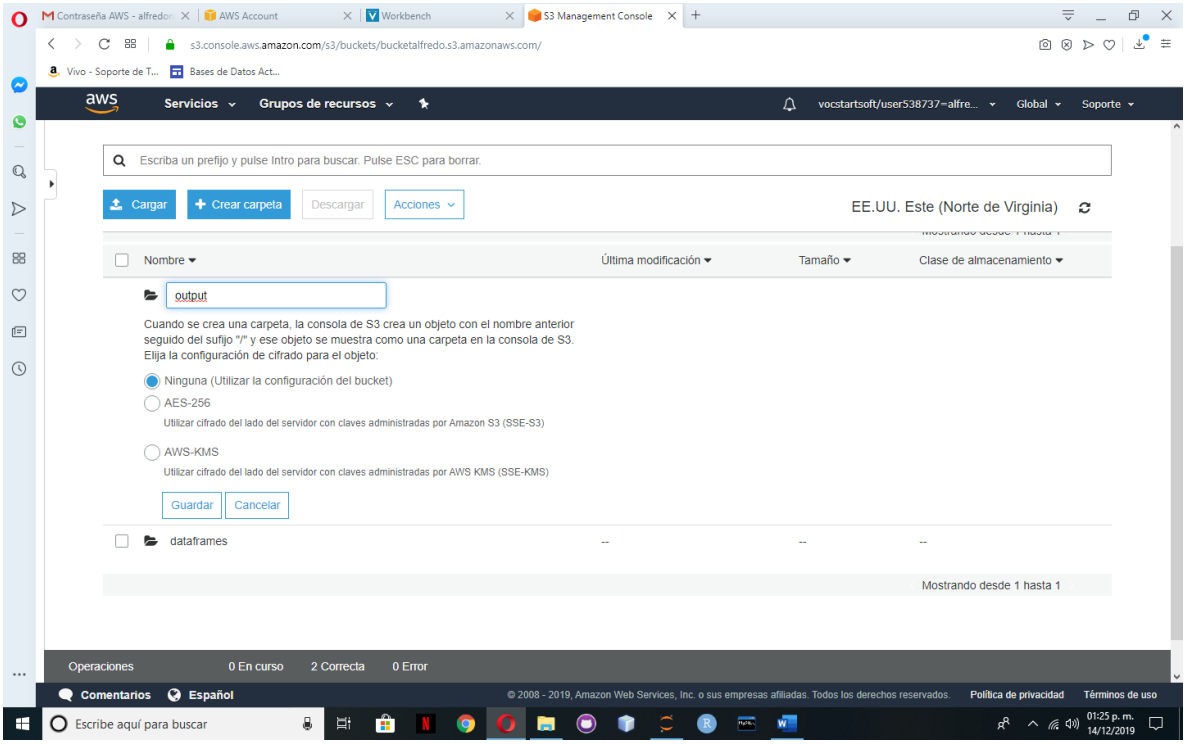
Usamos la convención para crear el bucket.



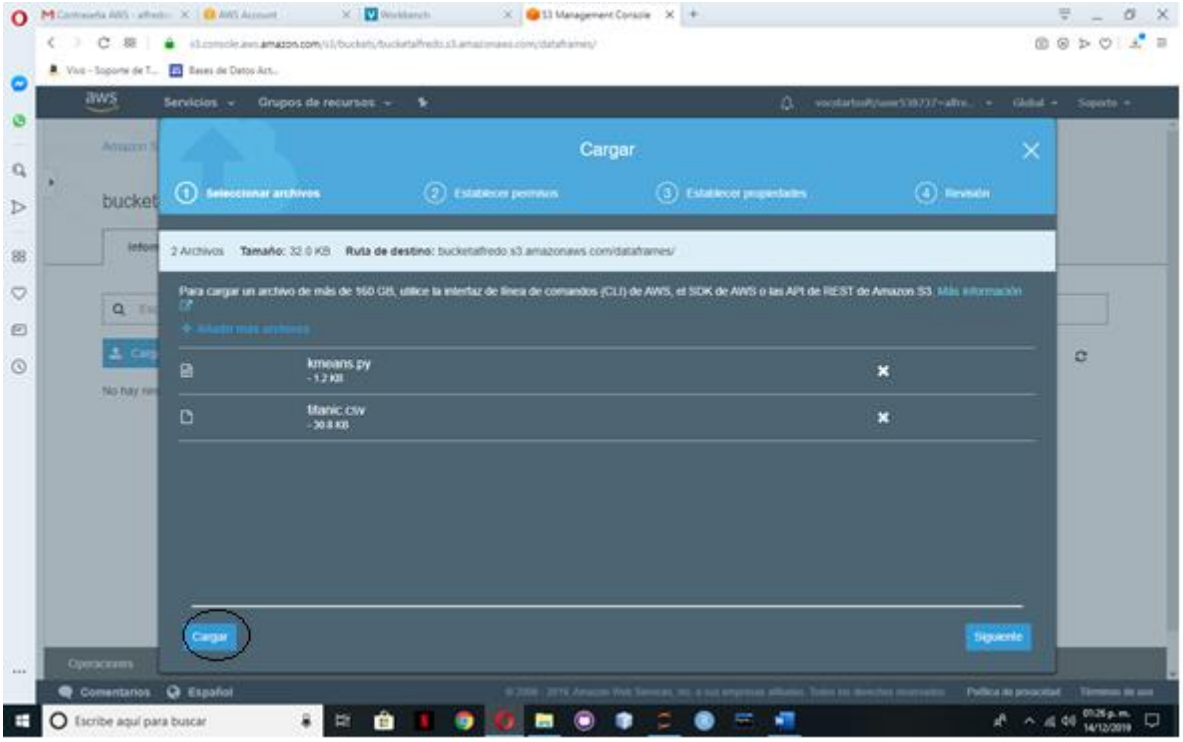
Se creó el bucket.

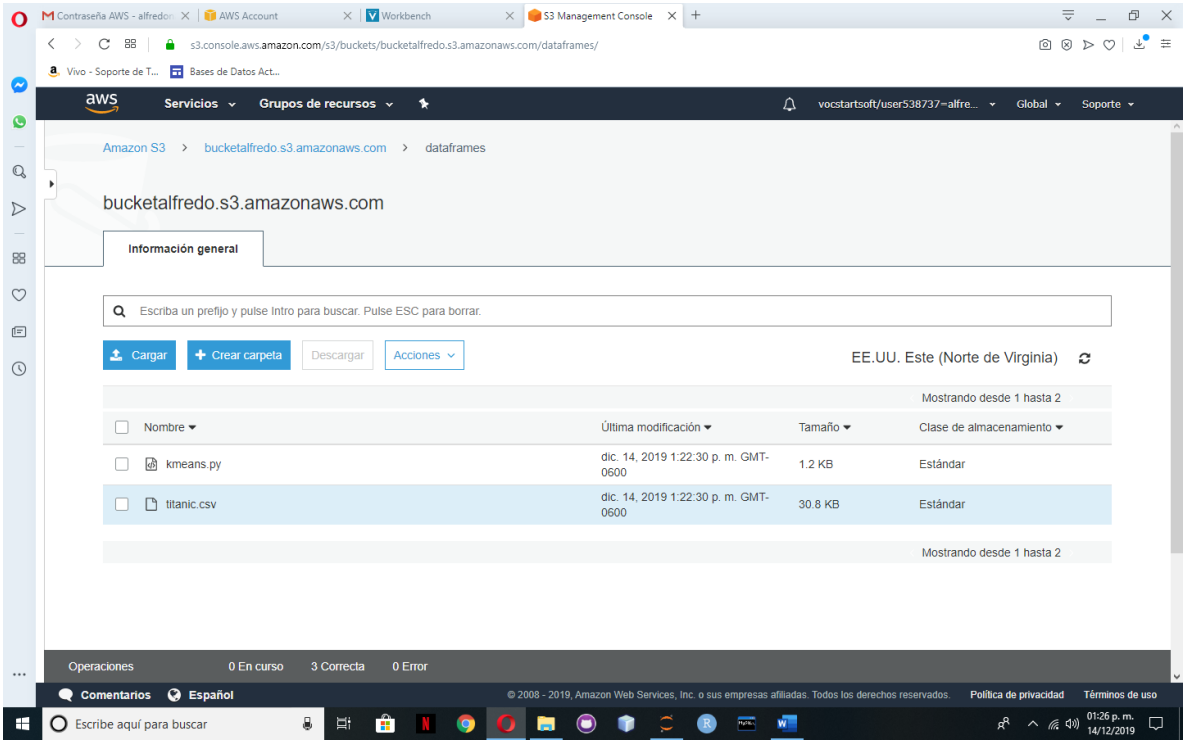


4.- Creamos las 2 carpetas que nos piden dataframes donde guardaremos los datos y output.



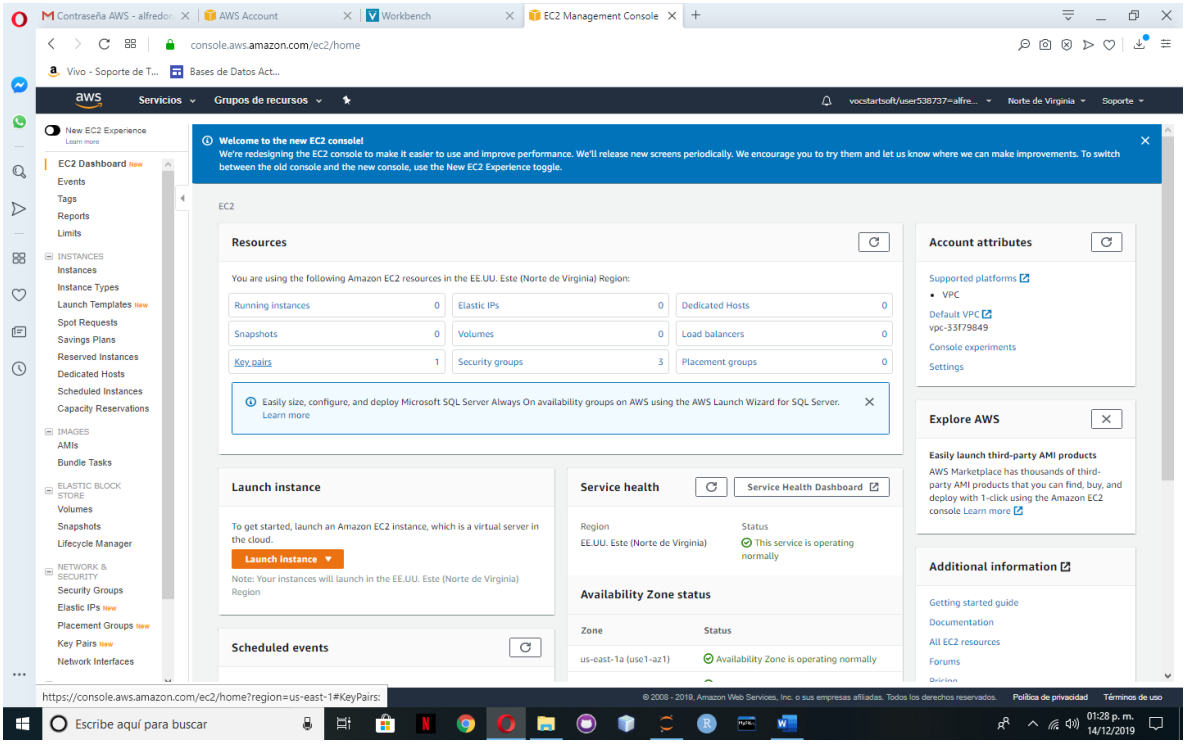
5.- Cargamos en la carpeta dataframes los archivos titanic.csv y kmeans.py.

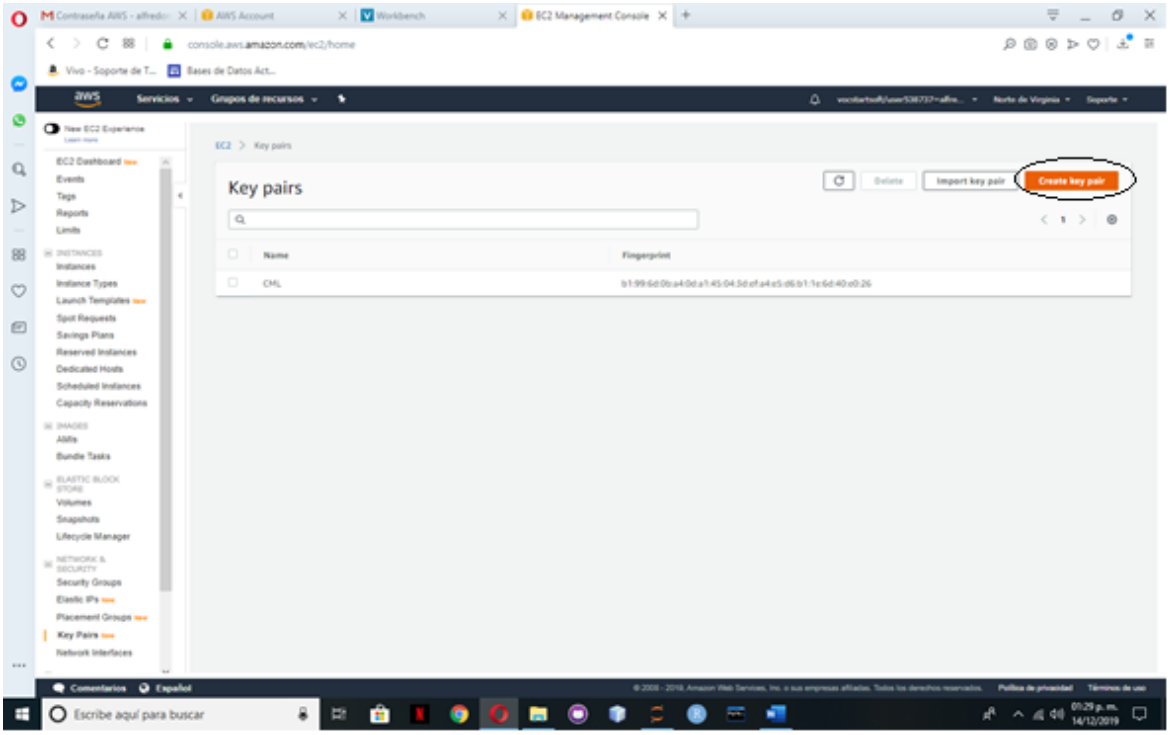




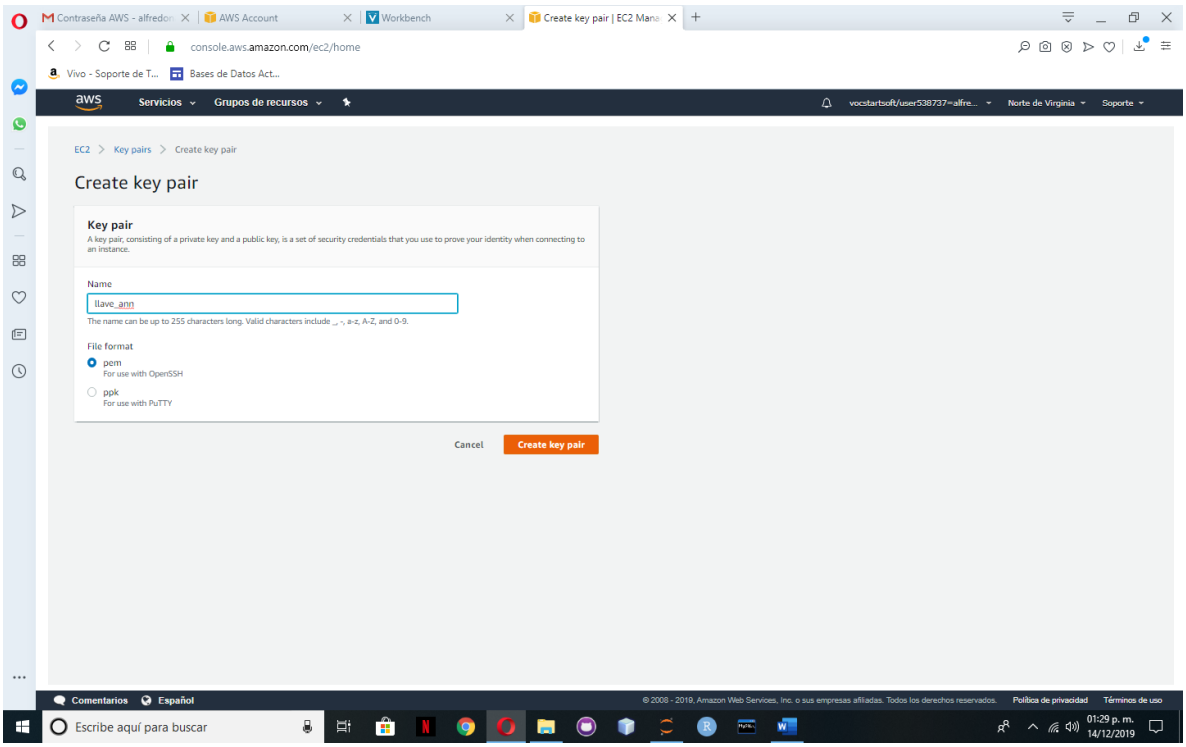
Pregunta 2

1.- Vamos a crear el par de claves EC2.

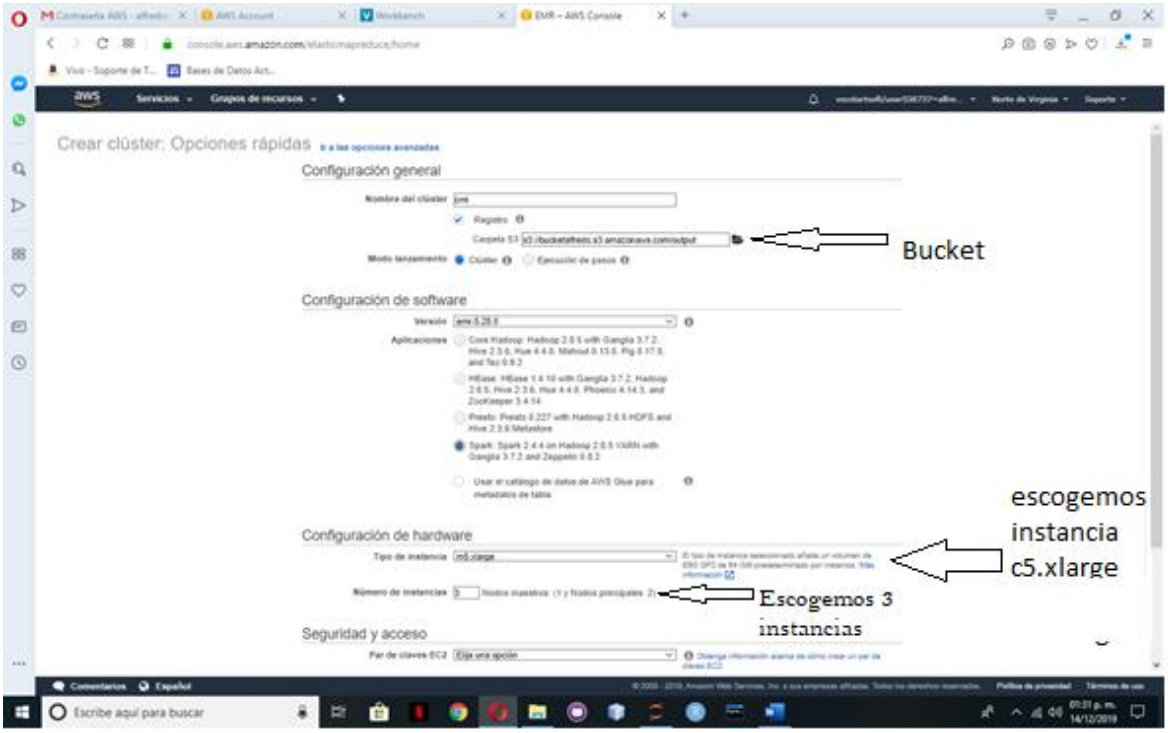


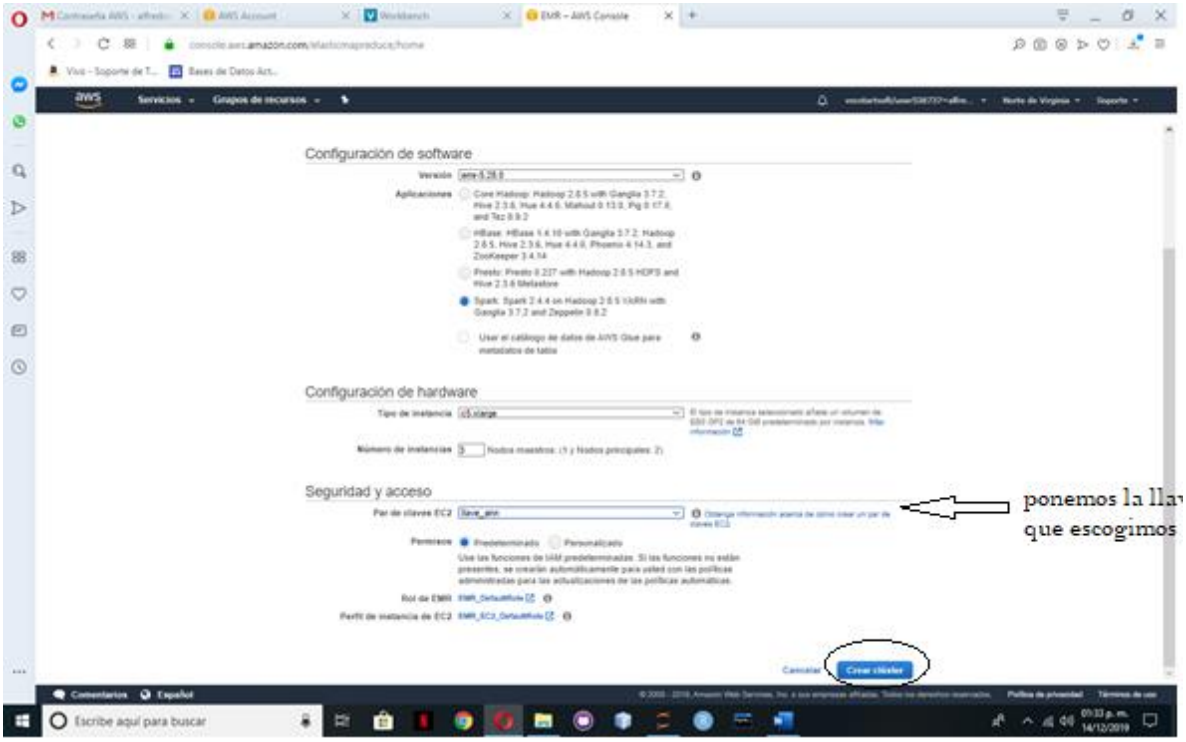


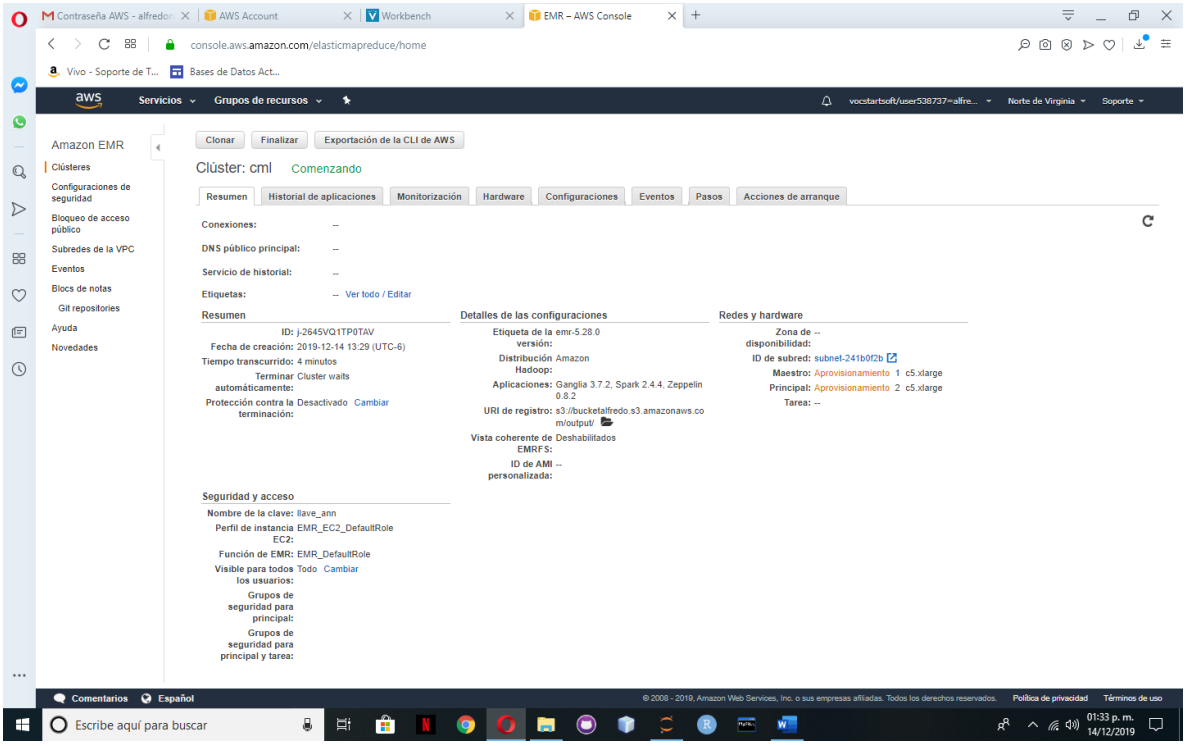
Elegimos formato .pem.

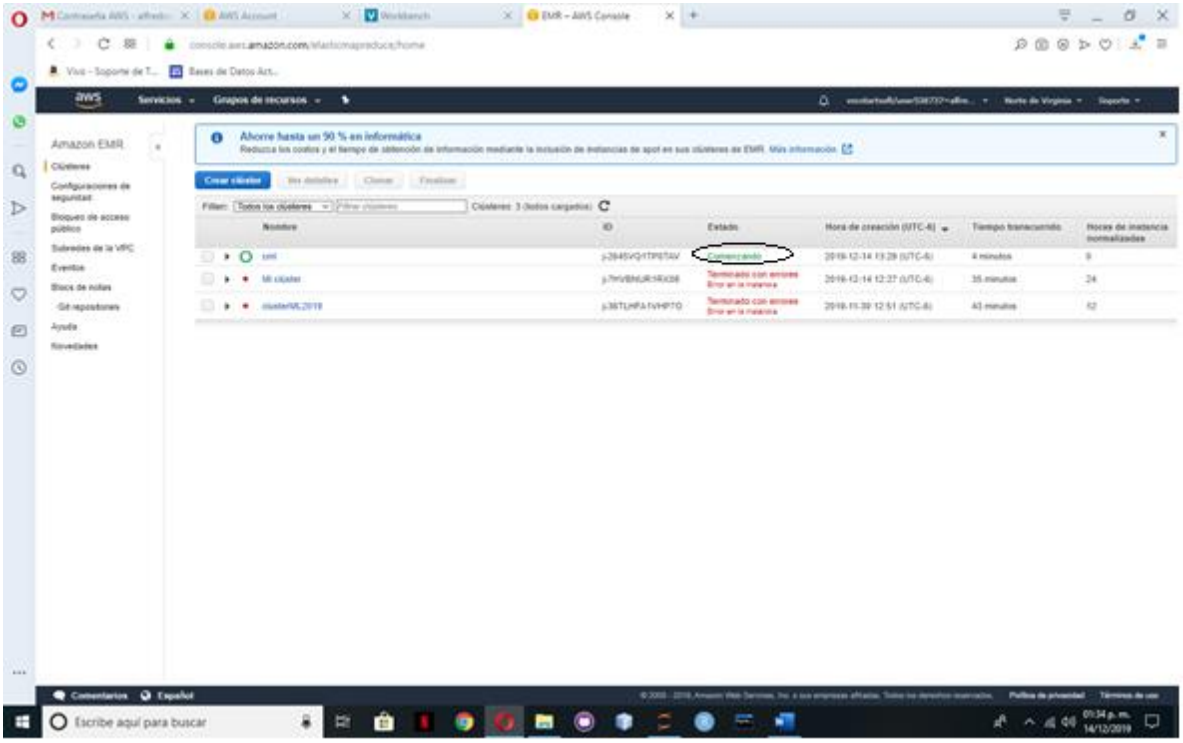


2.- Creamos el cúster en la consola de aws buscamos EMR.

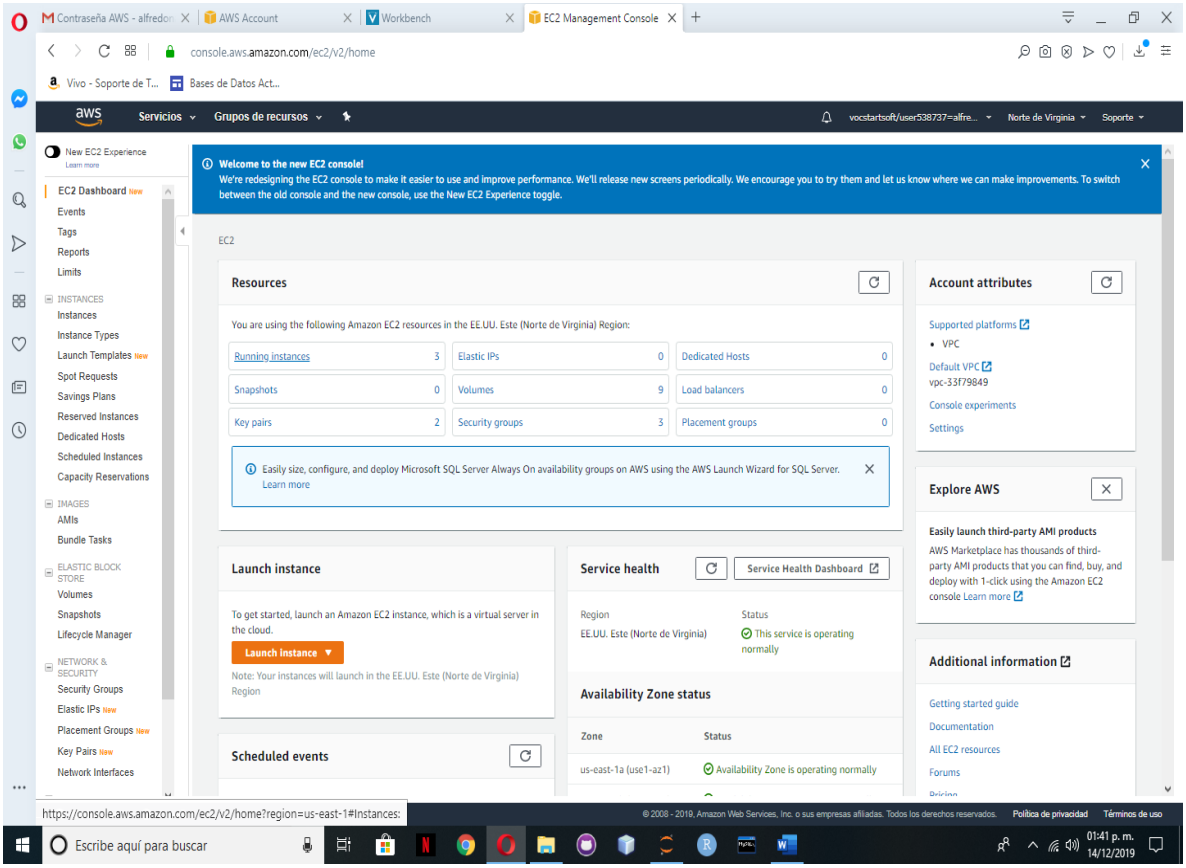


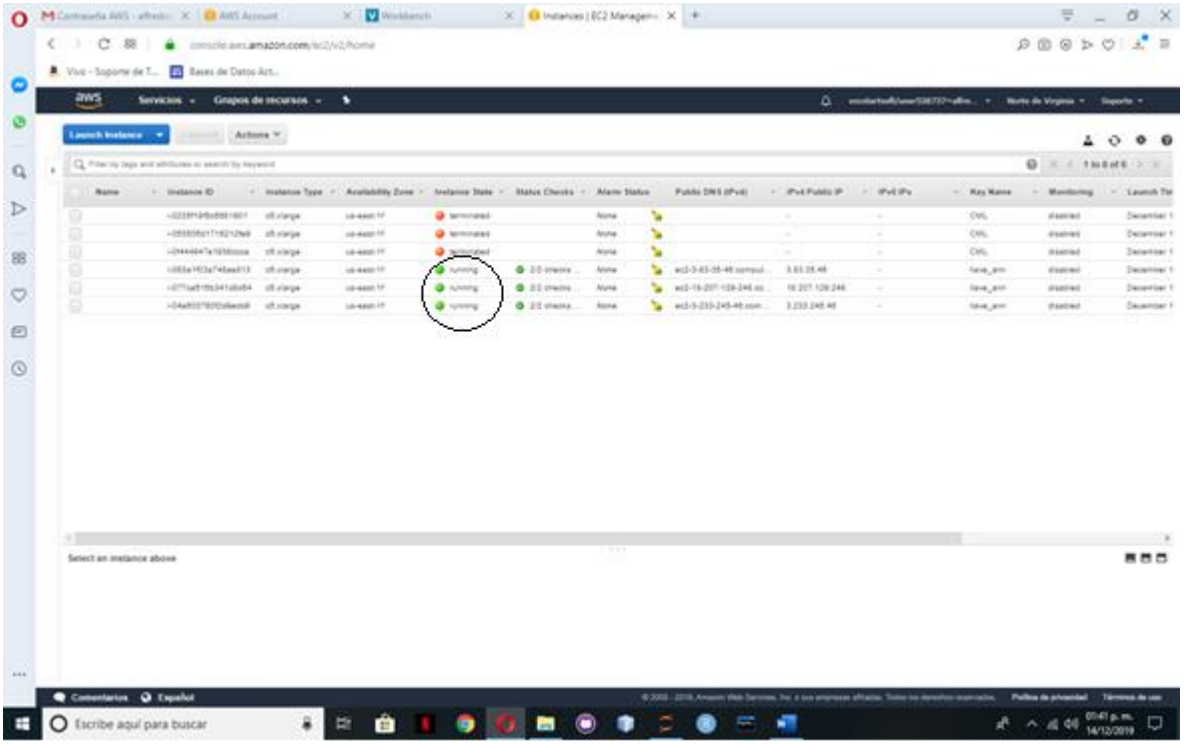

Configuración de software
Configuración de hardware
Seguridad y acceso
ponemos la llav
que escogimos
Crear clúster


EMR – AWS Console
console.aws.amazon.com/elasticmapreduce/home
Amazon EMR
Clústeres
Configuraciones de seguridad
Bloqueo de acceso público
Subredes de la VPC
Eventos
Blocs de notas
Git repositories
Ayuda
Novedades
Clonar
Finalizar
Exportación de la CLI de AWS
Clúster: cml Comenzando
Resumen
Historial de aplicaciones
Monitorización
Hardware
Configuraciones
Eventos
Pasos
Acciones de arranque
Conexiones: --
DNS público principal: --
Servicio de historial: --
Etiquetas: -- Ver todo / Editar
Resumen
ID: j-2645VQ1TP0TAV
Fecha de creación: 2019-12-14 13:29 (UTC-6)
Tiempo transcurrido: 4 minutos
Terminar automáticamente: Cluster waits
Protección contra la terminación: Desactivado Cambiar
Detalles de las configuraciones
Etiqueta de la versión: emr-5.28.0
Distribución Hadoop: Amazon
Aplicaciones: Ganglia 3.7.2, Spark 2.4.4, Zeppelin 0.8.2
URI de registro: s3://bucketalfredo.s3.amazonaws.com/output/
Vista coherente de EMRFS: Deshabilitados
ID de AMI personalizada: --
Redes y hardware
Zona de disponibilidad: --
ID de subred: subnet-241b0f2b
Maestro: Aprovisionamiento 1 c5.xlarge
Principal: Aprovisionamiento 2 c5.xlarge
Tarea: --
Seguridad y acceso
Nombre de la clave: llave_ann
Perfil de instancia EC2: EMR_EC2_DefaultRole
Función de EMR: EMR_DefaultRole
Visible para todos los usuarios: Todo Cambiar
Grupos de seguridad para principal:
Grupos de seguridad para principal y tarea:
Comentarios
Español
© 2008 - 2019, Amazon Web Services, Inc. o sus empresas afiliadas. Todos los derechos reservados.
Política de privacidad
Términos de uso
Escribe aquí para buscar
01:33 p. m.
14/12/2019

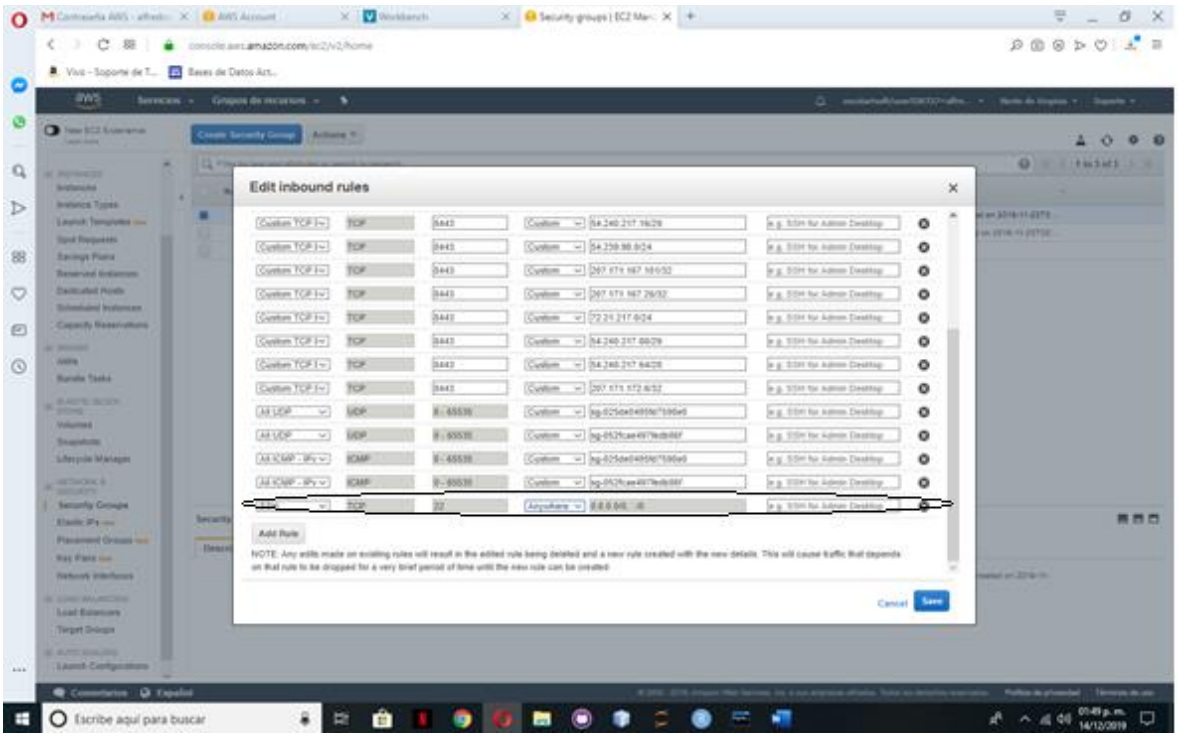


Veamos que nuestras instancias están corriendo.



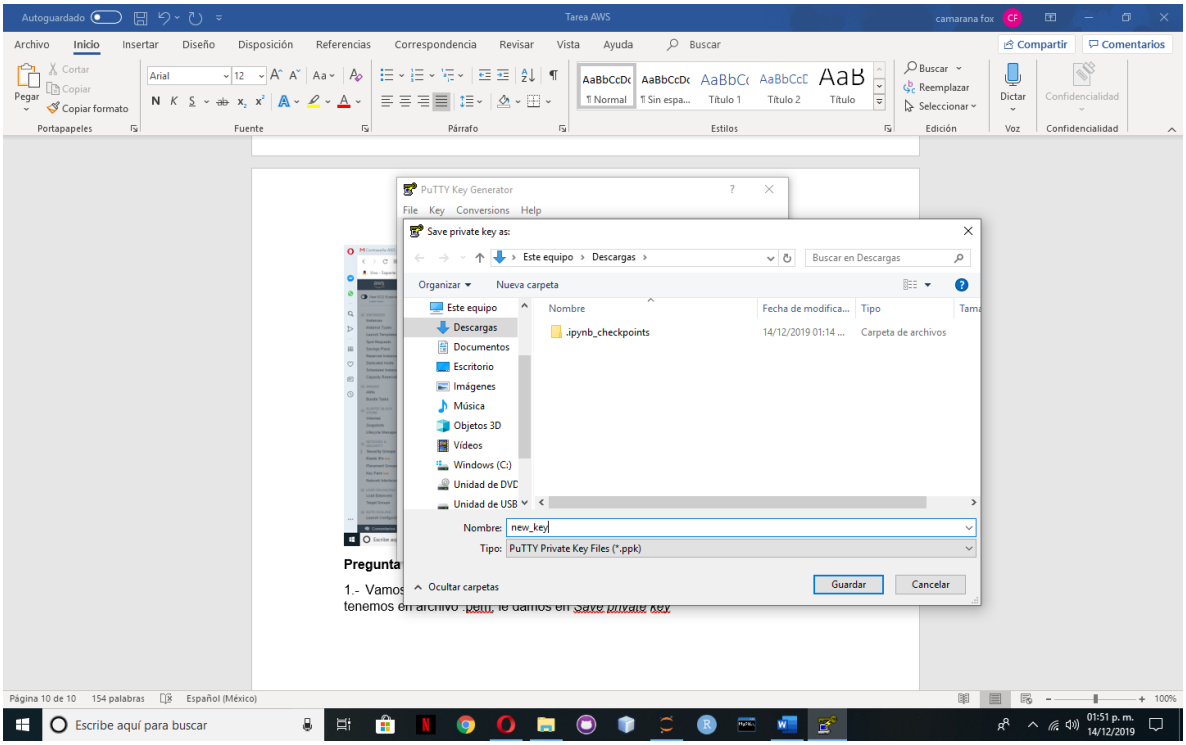


Mejoramos la seguridad agregando al máster una regla con las siguientes características (esta hasta el final).

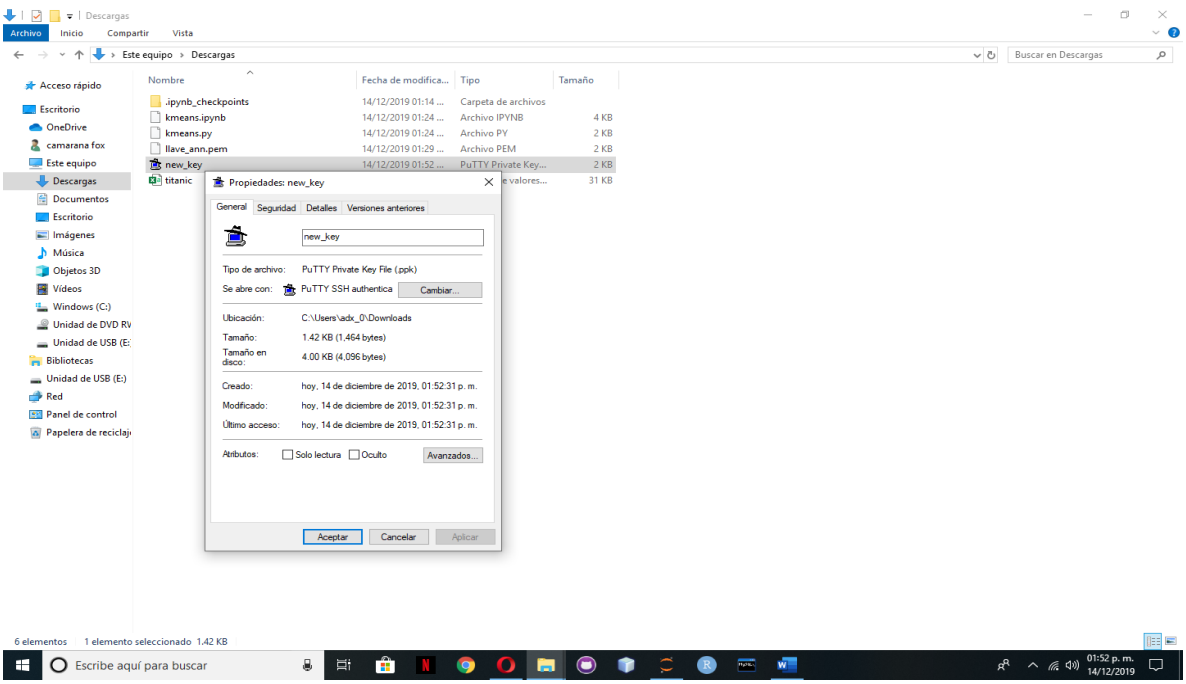


Pregunta 3

1.- Vamos a transformar la llave con *PUTTYgen* y transformamos la llave que tenemos en archivo .pem, le damos en *Save private key* y salvamos la llave bajo el nombre *new_key* con el formato *ppk.*

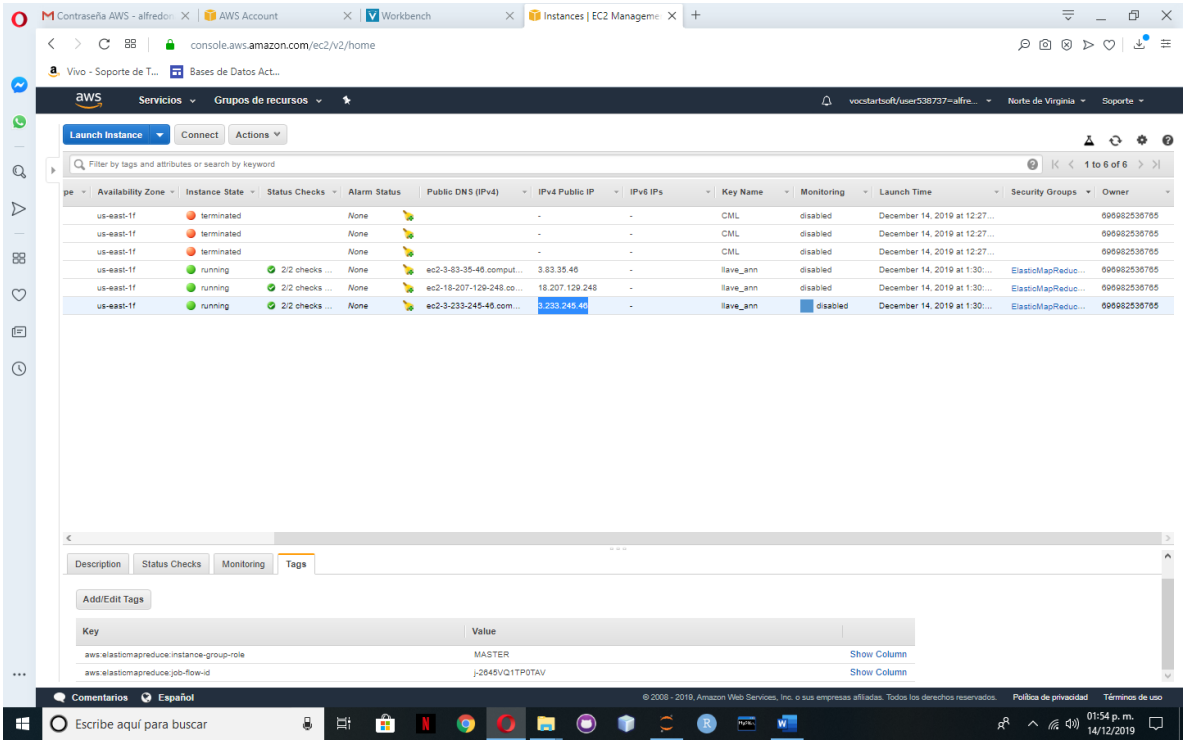


Veamos que esto fue así.

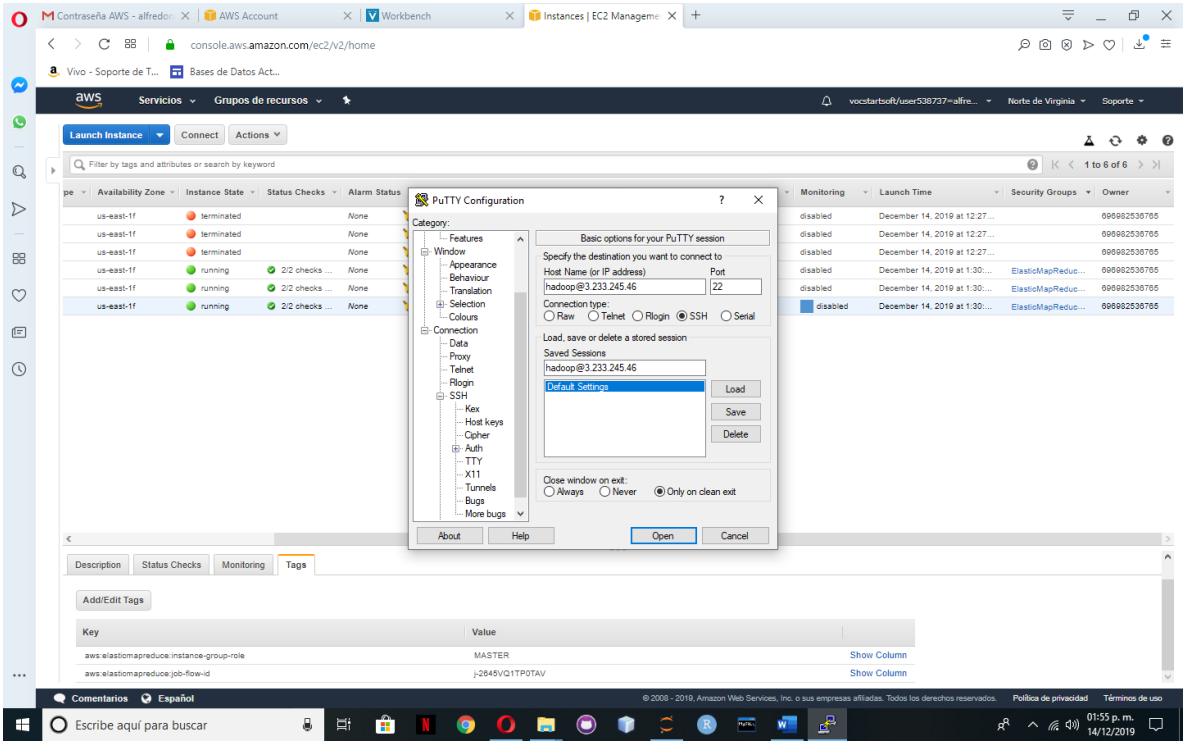


Pregunta 4

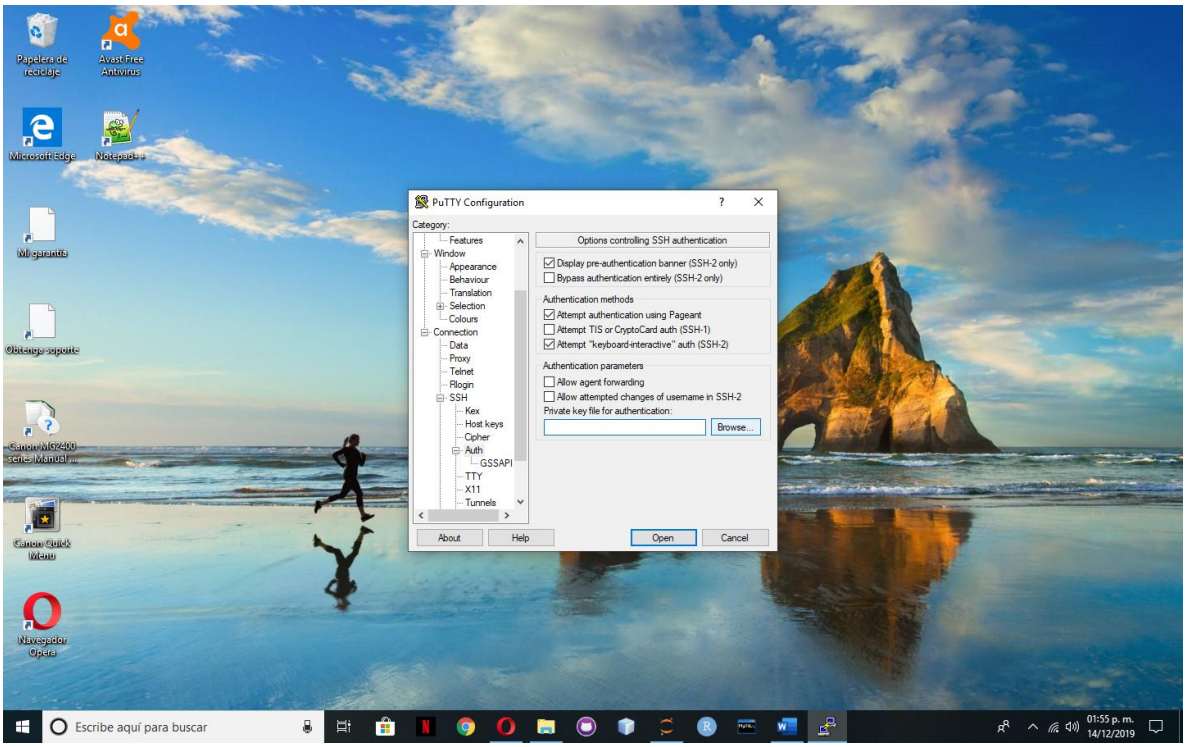
1.- Para conectarnos en el nodo maestro, copiamos su IP.



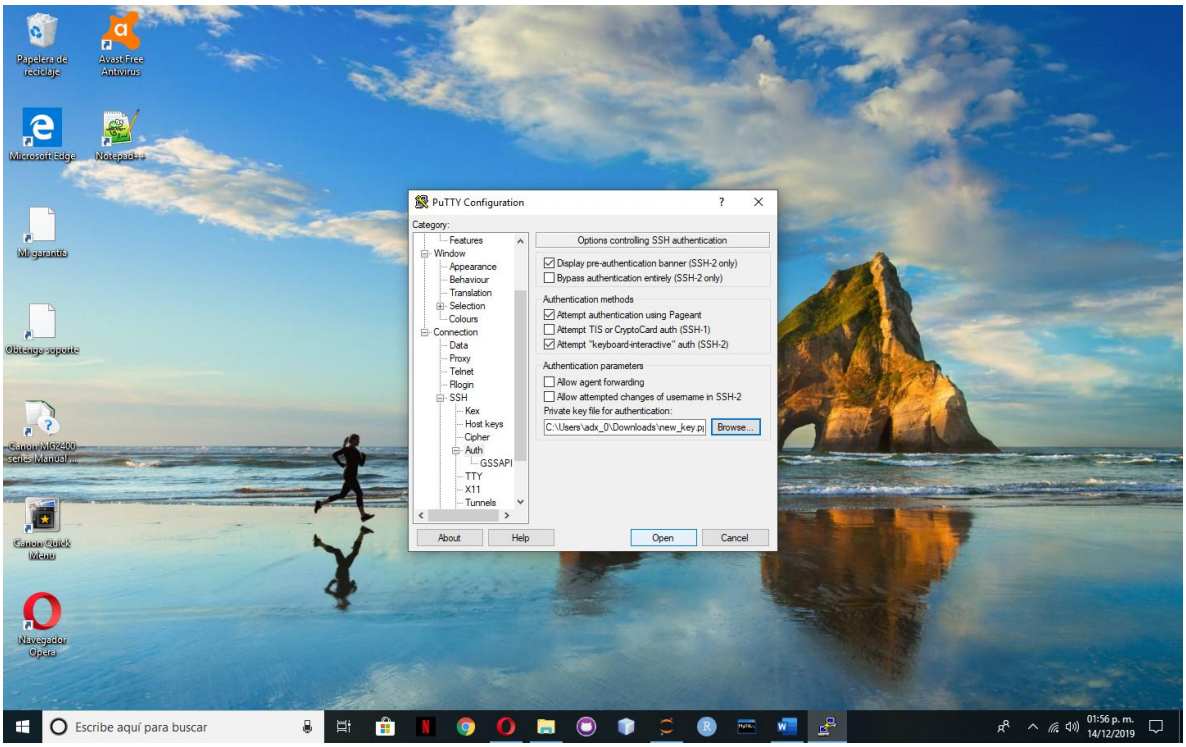
2.- Abrimos Putty y en la siguiente ventana configuramos.



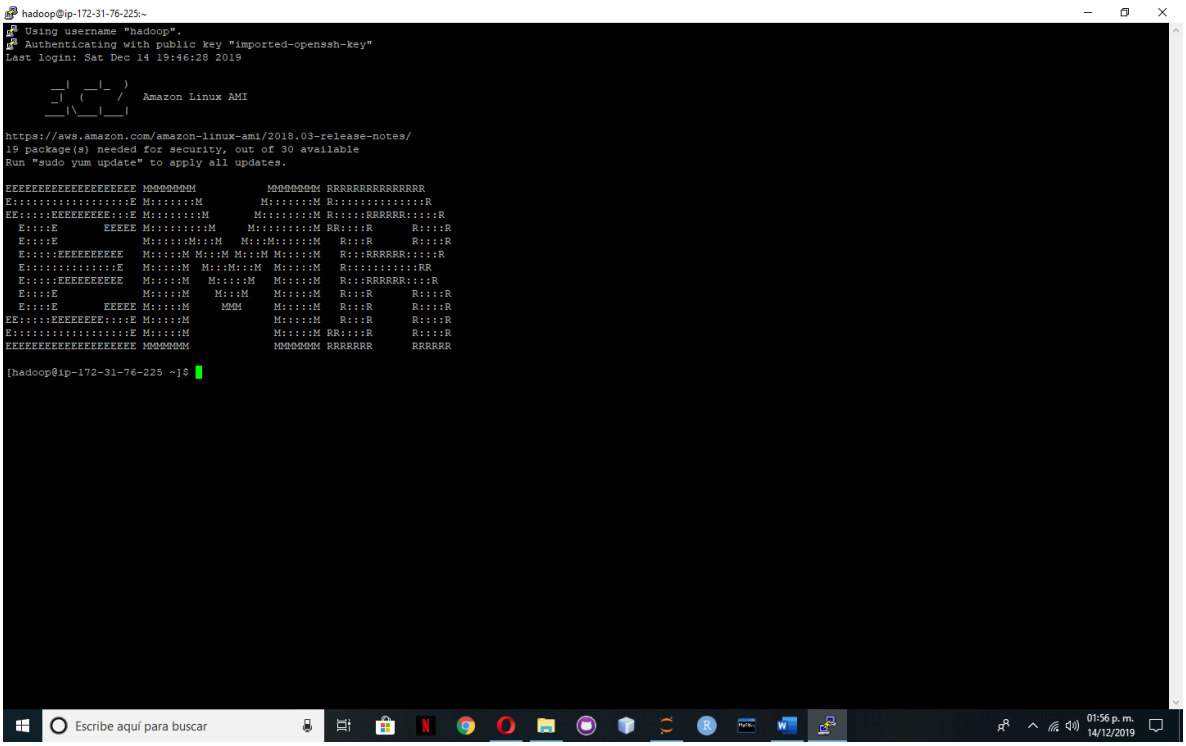
3.- Vamos a cargar nuestra llave.



Damos click en open.

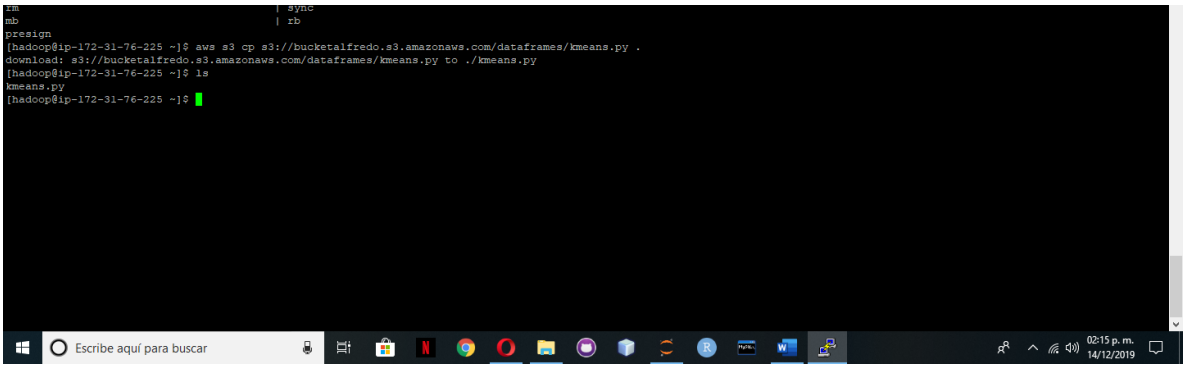


Podemos ver que todo es correcto porque nuestra instancia ya esta corriendo al aparecer el dibujo de EMR.

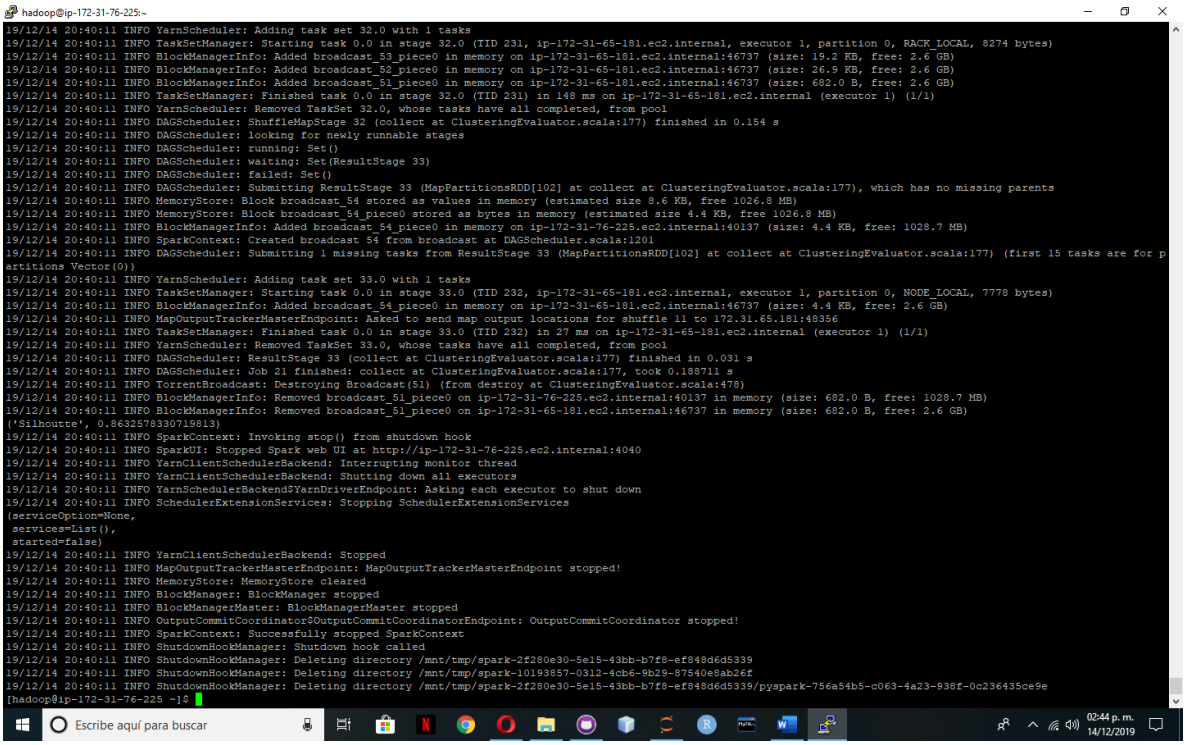


Pregunta 5

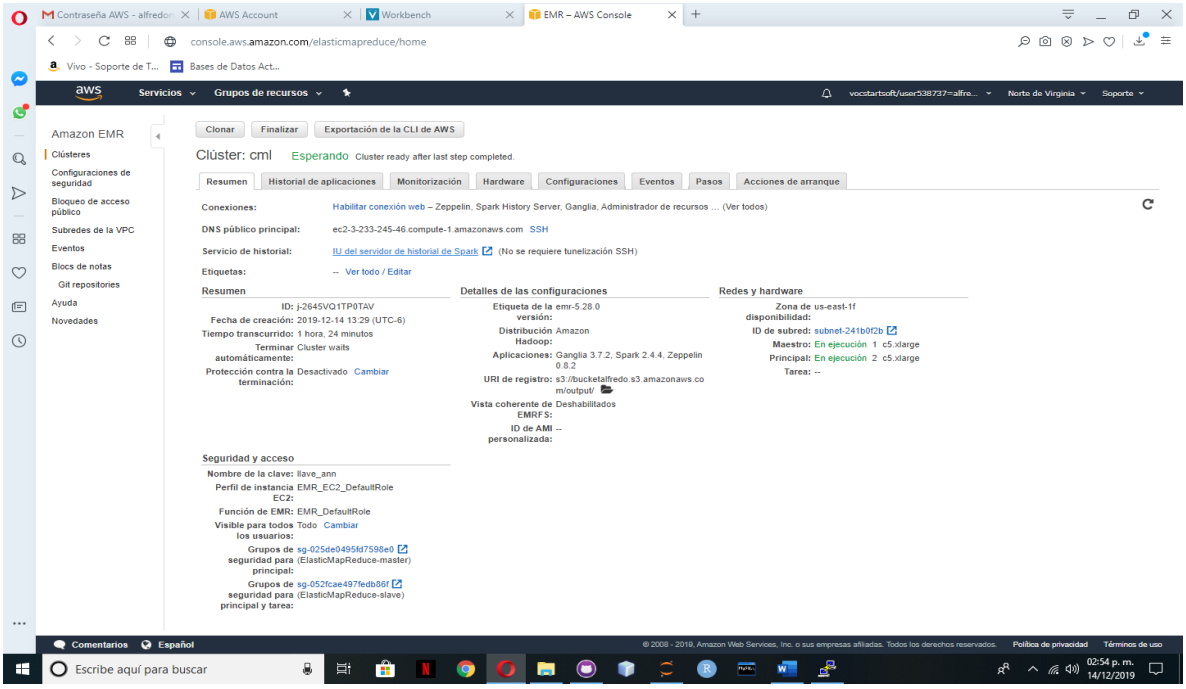
1.- Vamos a implementar los comandos dados en las instrucciones (copiamos el archivo del bucket) y aplicamos ls.

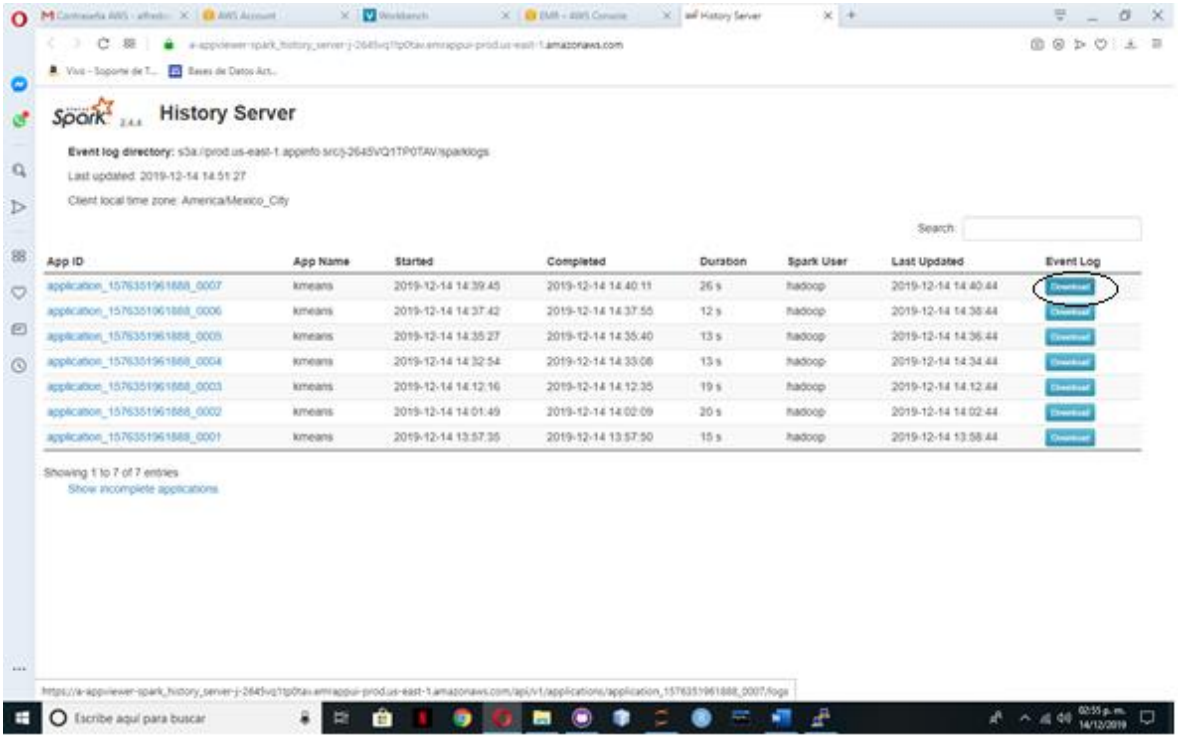


2.- Aplicamos el paso más importante que es spark.submi y vemos que todo esta bien porque nos sale el silohuette.

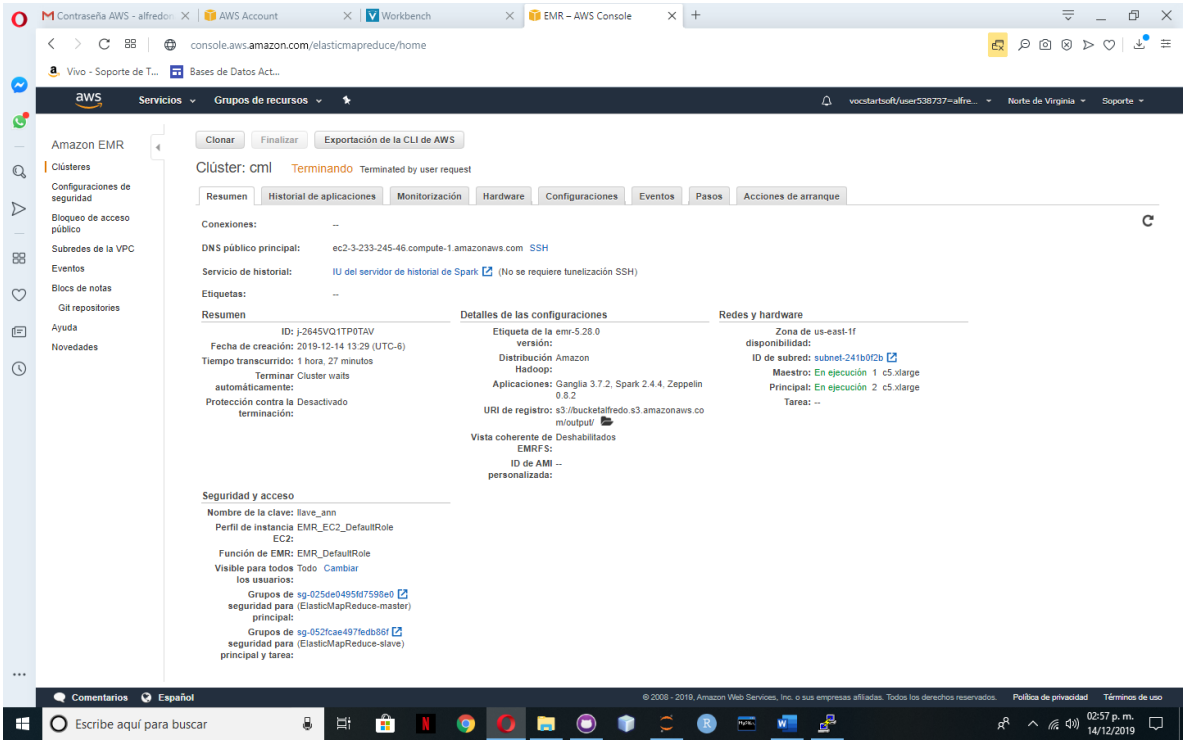


3.- Descargamos los logs que se van a adjuntar.

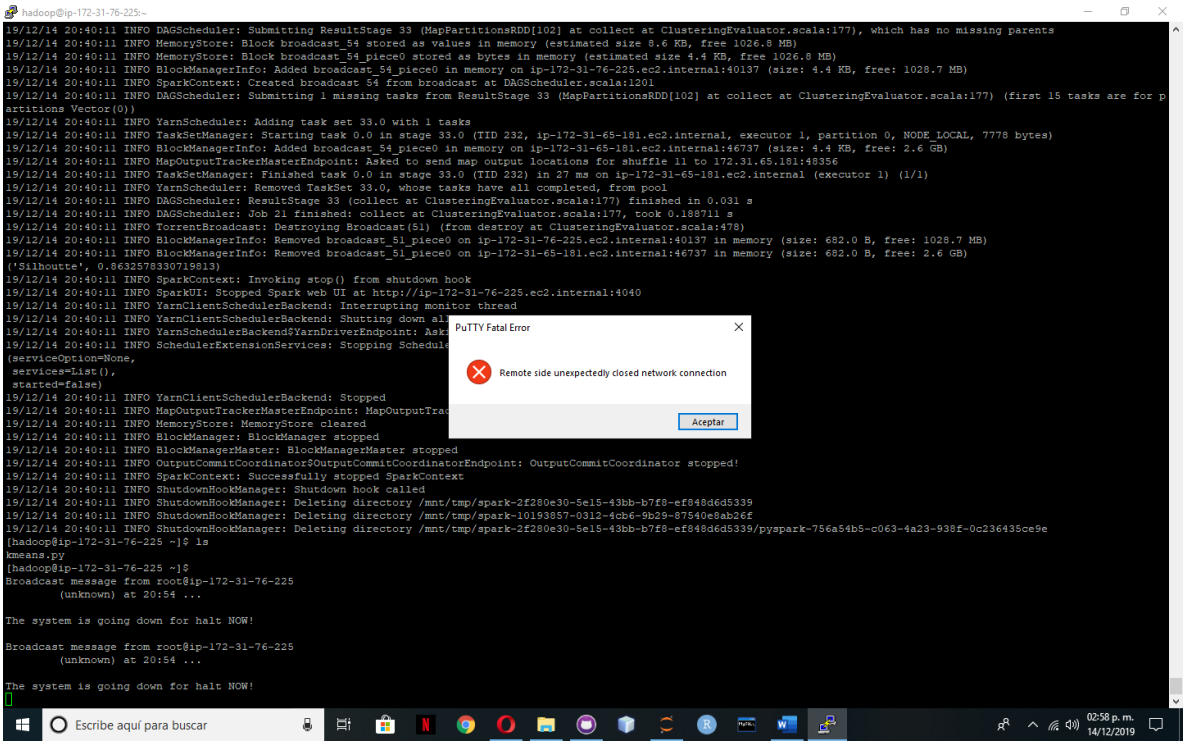


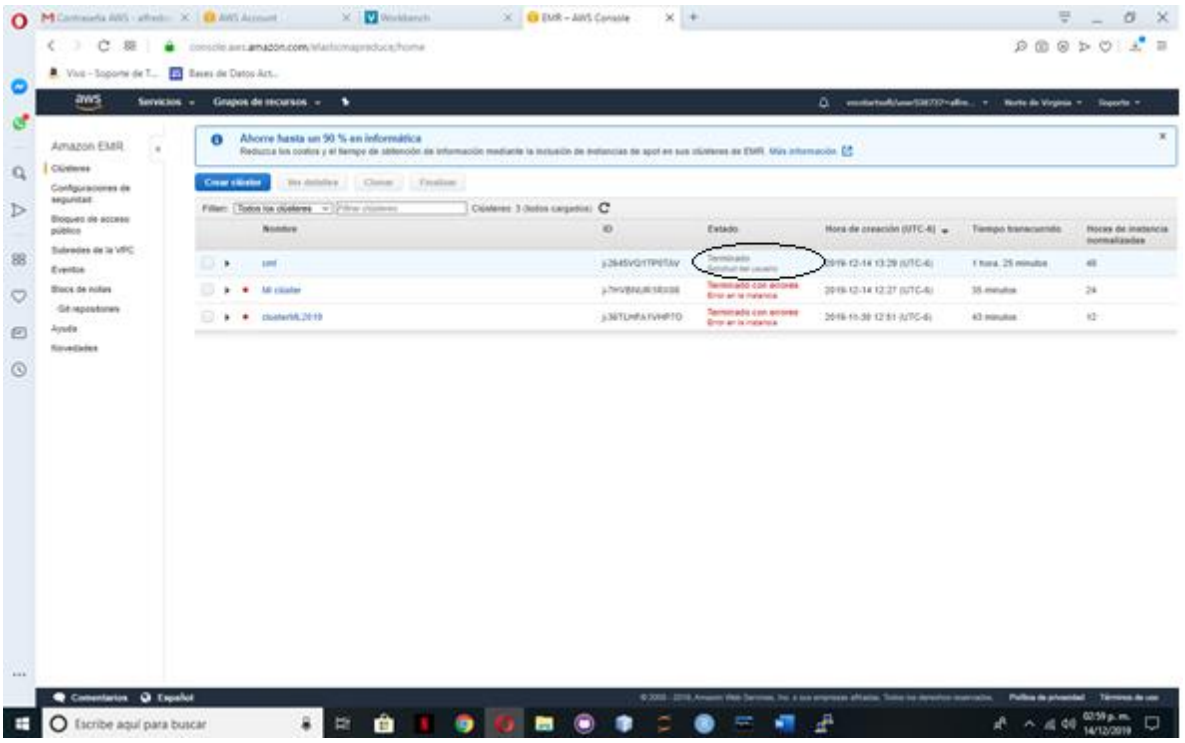


4.- Finalizamos el clúster para no gastar nuestros créditos.



Veamos que en Putty nos aparece en la consola un mensaje que es porque finalizamos el clúster.





6.- Se termina la práctica.