

Informe de Laboratorio 01

José André La Torre Esquivel

December 15, 2020

Introducción

Introducción a big data con Amazon EMR

I. OBJETIVOS

- Presentar el Informe.

II. DESARROLLO

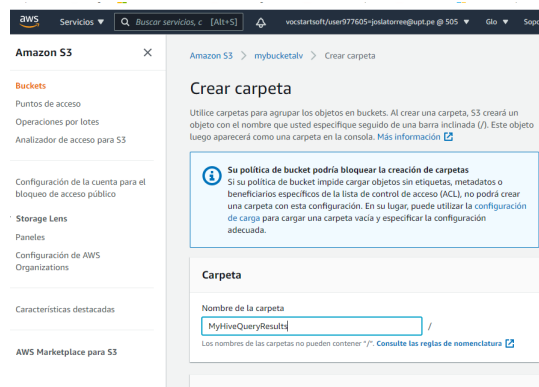


Figure 1: Creacion de Bucket en Amazon S3

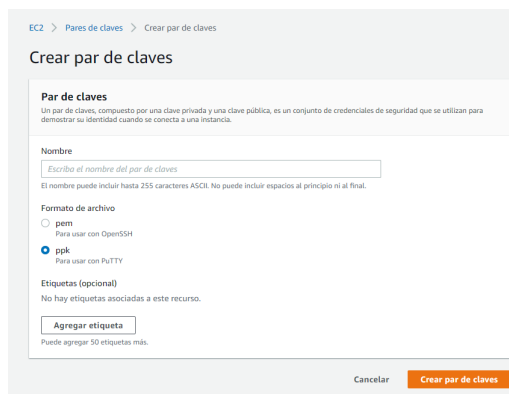


Figure 2: Creacion de Llaves de Amazon EC2

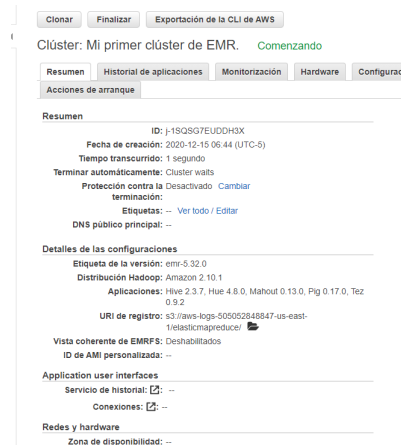


Figure 3: Lanzamiento y Creación de Cluster

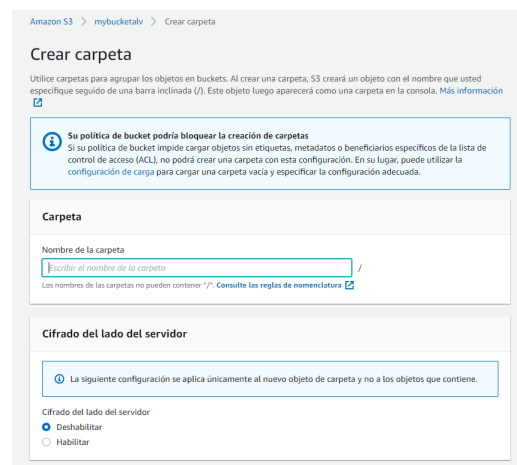


Figure 4: Lanzamiento y Creación de Cluster

Reglas de entrada				
Reglas de entrada				
Tipo	Protocolo	Intervalo de puertos	Origen	Descripción : opcional
Todos los TCP	TCP	0 - 65535	sg-01e96d3dad48470c6 (ElasticMapReduce-master)	-
Todos los TCP	TCP	0 - 65535	sg-0e6b500f7fecba2af (ElasticMapReduce-slave)	-
SSH	TCP	22	190.42.16.25/32	Nueva regla de 19

Figure 5: Eliminacion y Creacion de Reglas de Entrada para el EMR.

Añadir paso

Tipo de paso

Programa de Hive

Nombre

Programa de Hive

Ubicación S3 del script

s3://us-east-1-elasticmapreduce-samples/cloudfront/c

La ubicación S3 del script de Hive.

Ubicación S3 de entrada

s3://region-elasticmapreduce-samples

La ubicación S3 de los archivos de entrada de Hive.

Ubicación S3 de salida

s3://mybucket/MyHiveQueryResults/

La ubicación S3 de los archivos de entrada de Hive.

Argumentos

Especifique los argumentos opcionales para su script.

Acción sobre el error

Continuar

Qué hacer si se produce un error en el paso.

Cancelar

Añadir

Figure 6: Uso de datos de Hive, Creacion de Pasos.

Clúster: Mi clúster otra vez 2 Esperando Cluster ready after last step completed.				
Resumen	Historial de aplicaciones	Monitorización	Hardware	Configuraciones
Acciones de arranque				
Concurrencia: 1 Cambiar				
After last step completes: Cluster waits				
<div>Añadir paso</div> <div>Clonar paso</div> <div>Cancelar paso</div>				
<div>Filter: Todos los pasos</div> <div>Pasos: 2 (todos cargados)</div>				
ID	Nombre	Estado	Hora de inicio (UTC-5)	Ti
s-EBARSYRRRX9Q	Programa de Hive	Completado	2020-12-15 08:22 (UTC-5)	34
Ubicación de JAR: command-runner.jar Clase principal: Ninguno Argumentos: hive-script --run-hive-script --args -f s3://us-east-1-elasticmapreduce-samples/cloudfront/c Acción sobre el error: Continuar				
s-GUSNZSJ40NB	Configurar depuración de Hadoop	Completado	2020-12-15 08:12 (UTC-5)	4

Figure 8: Corre adecuadamente y aparece como completado, el resto es terminacion de Cluster y eliminacion de Bucket.

Amazon S3 > aws-logs-50505284847-us-east-1 > elasticmapreduce-samples/ > cloudfront/

cloudfront/

Información general sobre carpetas

Región

EE. UU. Este (Norte de Virginia) us-east-1

URI DE S3

s3://aws-logs-50505284847-us-east-1-elasticmapreduce-samples/clo

Arrastre y suelte los archivos y las

Objetos (1)

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Para que otras personas obtengan acceso a los objetos

Eliminar

Acciones

Crear carpeta

Cargar

Buscar objetos por prefijo

Nombre

Tipo

Última modificación

data_folder\$

-

15 Dec 2020 8:22:58 AM -05

Figure 7: Primer error de creacion, vuelvo a intentar.

2

III. CONCLUSIONES

Fue complicado realizarlo rapido, especialmente por no considerar el lugar de creacion, "Eu-West-2".

IV. RECOMENDACIONES

Tiempo para analizar el posible error.