

Esquema de desarrollo en la nube

Luna Duran – Taller 7

Configuración en AWS:

1. Abrir AWS
2. Ingresar a **EMR**
3. Crear un *cluster* con las siguientes características:
 - a. Este configurado en *Spark*
 - b. Como grupos de instancia tenga:
 - i. 1 nodo primario de tipo EC2 *m4.large*
 - ii. 1 nodo *core* de tipo EC2 *m4.xlarge*
 - iii. 2 nodos *task* de tipo EC2 *m4.xlarge*
 - c. Configuración del tiempo de terminación del *cluster* en 1 hora
 - d. En configuraciones de seguridad seleccionar una *key pair* para el **SSH** del cluster
 - i. Para esto vamos los servicios de **AWS**
 - ii. Ingresamos a **EC2**
 - iii. Seleccionamos *Grupos de seguridad*
 - iv. Seleccionamos *key Pairs*
 - v. Creamos un nuevo par, con las configuraciones:
 1. Formato de llave *privada* *.ppk* (si en Windows que fue mi caso)
 - vi. Luego de creada la llave se descargará en el computador, la guardamos en la carpeta donde están los archivos de código
 - e. En los roles de servicio **EMR** escogemos *EMR_DefaultRole*
 - f. En el perfil de la instancia **EC2** para Amazon EMR escogemos *EMR_EC2_DefaultRole*
 - g. Creamos el *Cluster* y esperamos a que cargue

Instalación de programas necesarios:

4. Descargue **Putty.exe**
5. Descargue **pscp.exe**

Carga de archivos a la nube

6. Abrimos la consola del computador en la carpeta donde tenemos nuestro archivo de código *.py* y nuestra llave *.ppk* y corremos el siguiente comando:

pscp -i nombre_llave.ppk parcial.py nombre_SSH:

- a. Donde *nombre_llave.ppk* es la llave descargada anteriormente, *parcial.py* es nuestro ejecutable de **Spark**, y *nombre_SSH* es el encontrado al seleccionar en el cluster la opción

Connect to the primary node using SSH

Conexión por medio de Putty

7. Ejecutamos **Putty.exe**

- a. Ingrese en la categoría de Sesión
 - i. En el nombre de host o IP address ingrese *nombre_SSH*
- b. Ingrese a la categoría de *Conexión*
- c. Ingrese a la categoría de *SSH*
- d. Ingrese a la categoría de *Auth*
- e. Ingrese a la categoría de *Credenciales*
 - i. En la llave privada para autenticación busque el archivo *nombre_llave.ppk*
- f. Seleccione el botón **Open** en la parte inferior
- g. En la ventana que aparece seleccione **aceptar**
- h. Se encuentra en la consola, para ejecutar el código escriba:

spark-submit parcial.py