Interfaz de usuario gráfica, Texto, Aplicación, Sitio web

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica, Sitio web

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Word

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Word

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Word

Descripción generada automáticamente

Cambio se políticas para el bucket para que sea público:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

* Query inicial en Raw con Athena

Este query trae el número total de personas vacunadas y de personas vacunadas por cada 100 habitantes por país/región y lo ordena en orden descendente, según el número de personas vacunadas por cada 100 habitantes. Es importante destacar que estos números de vacunados son acumulativos en las tablas originales. Esto quiere decir que en cada fecha, se reporta el número total de vacunados a la fecha, no el número de vacunados nuevos en esa fecha, lo cual quiere decir que en cada fecha, los vacunados son los de la fecha anterior más los que se vacunaron en esa fecha y, por este motivo, se toma la cifra máxima de vacunados en la tabla y no la suma ni la cuenta.

Interfaz de usuario gráfica

Descripción generada automáticamente

* Glue ETL

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Se quitaron todos los espacios y las tildes en esta transformación

Diagrama

Descripción generada automáticamente

Se guardan los datos transformados en la zona “Trusted”

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Propiedades del Job

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Se agregó AmazonS3FullAccess al rol IAM de Glue

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Corrida del Job exitosa

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Nuevo crawler para la zona Trusted.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Resultado después del proceso de ETL con Glue y el crawler en trusted: ya las columnas están bien nombradas.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Query en Athena usando los nuevos nombres de columnas

**Creación del EMR**

Elegir las herramientas adecuadas para el cluster. (Nota: en este caso no se usarán todas, pero igualmente es útil porque quedará como plantilla para trabajos futuros en los que sí se usen Spark, notebooks, etc.)Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Se escogen los tipos de nodos, que en este caso serán m4.xlarge y se pondrán en Spot para que salga más barato.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* Nombre del cluster

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

* Key-pair del cluster
* Interfaz de usuario gráfica, Texto, Aplicación

  Descripción generada automáticamente
* Se debe abrir el puerto 8888 para tener acceso a Hue para hacer las consultas de Hive.

Tabla

Descripción generada automáticamente

* En security groups, se le cambiarán las inbound rules al nodo maestro del EMR para abrir el puerto 8888 y permitir acceso desde cualquier dirección IP.

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

* Ya se tiene acceso a la interfaz gráfica de Hue y se ven las bases de datos creadas por los crawlers.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

* Ejemplos de consultas en Hive, tanto en la base de datos original como en la modificada en trusted.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Esta consulta en Hive muestra el número total de muertes por país en orden descendente.

Una captura de pantalla de una computadora

Descripción generada automáticamente

Esta consulta en Hive muestra el número total de casos por cada tipo de contagio. La consulta fue hecha sobre los datos en la zona Trusted, que se obtuvieron tras quitar los espacios en los nombres de las columnas.