|  |
| --- |
| Programa AWS Big Data Analytics Specialist |
| Informe  Arquitectura Final AWS  Integrantes:  Anthony Steve Brian Manosalva López  Julian Castiblanco P  Jennyfer Nereida Combariza Vanegas  Isel Paola Santos Rojas |
| Junio 2020 |

|  |
| --- |
|  |

# Introducción

Este proyecto constituye el informe final del curso AWS Big Data Analytics Specialist.

Comprender las emociones de las personas cada día se convierte en una de las herramientas más esenciales para las empresas, ya que les permite escuchar atentamente a sus clientes y adaptar sus productos y/o servicios para satisfacer sus necesidades.

En este breve proyecto vamos a explicar como hacer una arquitectura serverless para realizar análisis de sentimientos de tweets relacionados con el covid en Perú.

Para esa arquitectura serverless usaremos los siguientes servicios de AWS:

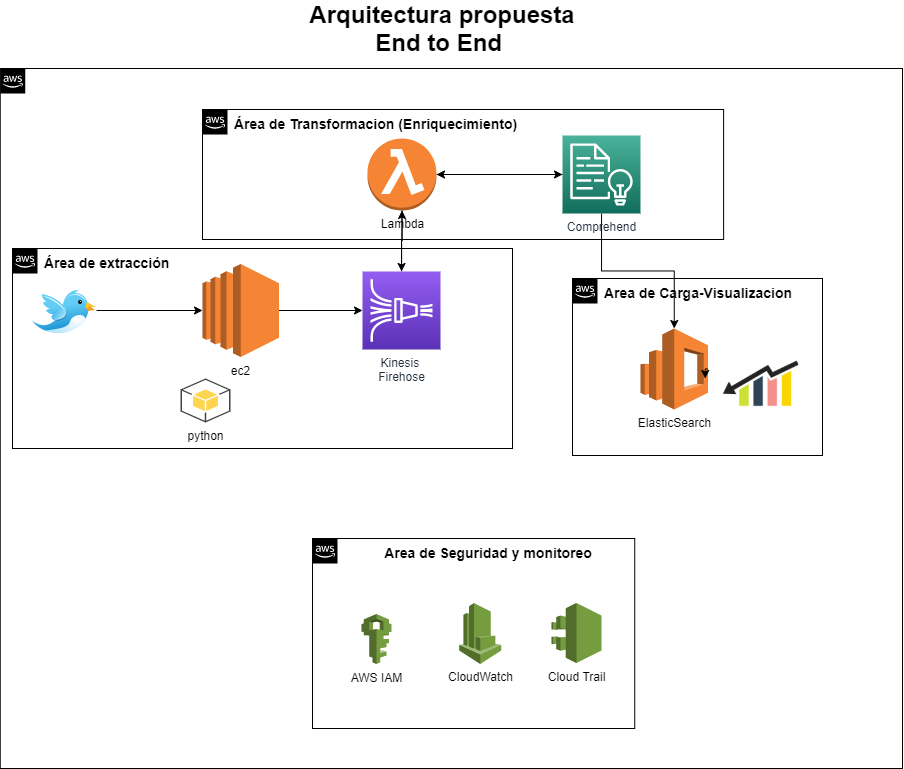
1. Kinesis Firehose: servicio de AWs que permite el procesar datos en tiempo real.
2. Lambda: servicio de AWS que ejecuta cómputo a demanda.
3. Comprehend: servicio de procesamiento de lenguaje natural.
4. Lambda: servicio de AWS que ejecuta cómputo a demanda.
5. Elastic Search: servicio de AWS facilita la implementación, la protección y la ejecución de Elasticsearch a escala de manera rentable
6. Kibana. Ventana al Elastic Stack. Kibana es una interfaz de usuario gratuita y abierta que te permite visualizar los datos de Elasticsearch

Para este análisis de sentimientos recopilamos los tweets con Kinesis Firehose, los procesamos con comprehend y visualizamos utilizando Elastic Search- Kibana.

# Objetivo

Construir una arquitectura serverless que permita la visualización de análisis de sentimientos de tweets en tiempo real.

# Arquitectura propuesta



# Alcance

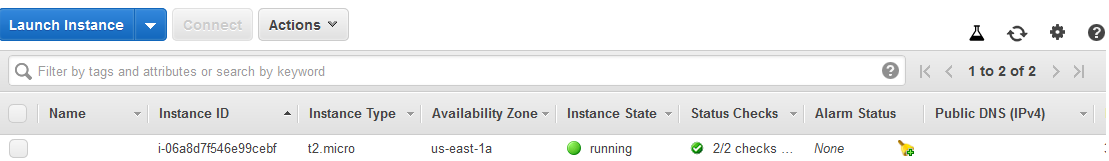
El presente documento describe los lineamientos llevados a cabo para la generación de la arquitectura propuesta desde la extracción de los tweets hasta la visualización de los resultados obtenidos del análisis de sentimientos.

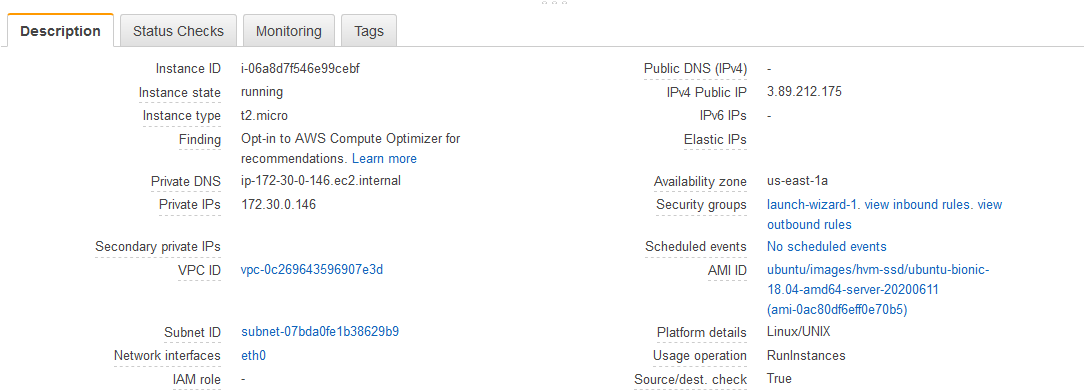
# Configuraciones generales

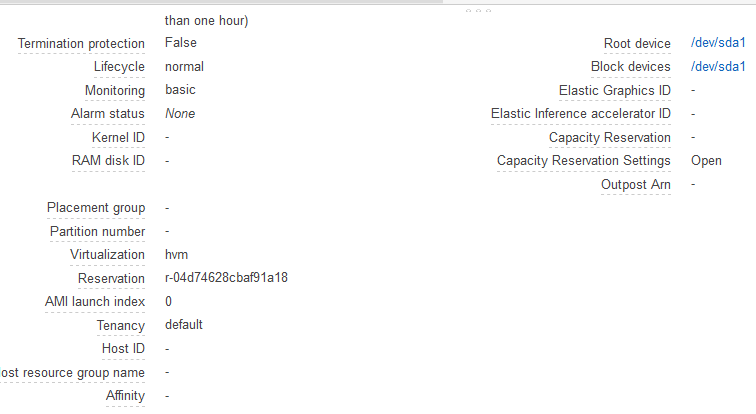
A continuación se listan los servicios que se usaron en el proceso de la generación de la arquitectura:

* 1. **Configuración de EC2**

Se crea una máquina EC2 que se encargará de ejecutar un código Python para conectarse con la cuenta de desarrollo de twitter y de esa manera extraer información acerca del covid en Lima Perú, con filtro de idioma en español.



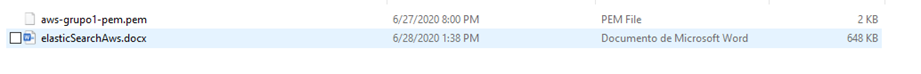




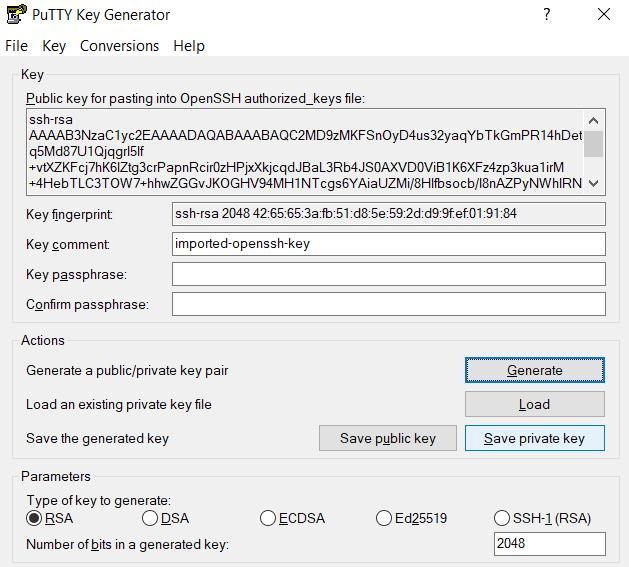
* 1. **Conexión a EC2 por SSH a través de Putty**

Objetivo: Cargar el código de python e instalar las librerías necesarias para una conexión exitosa.

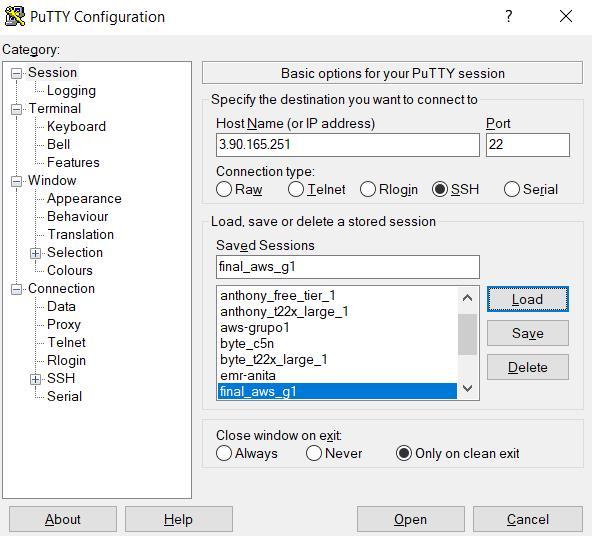
* Paso 1: Se descarga el archivo pem, donde viene la key privada.



* Paso 2: Se abre la herramienta “putty generator” para convertir el archivo pem en ppk que es válido para la conexión desde putty.



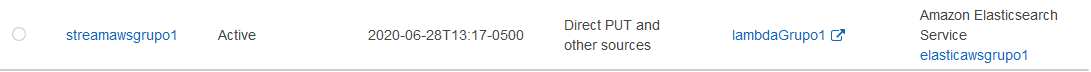
* Paso 3: Se hace conexión a través de putty.

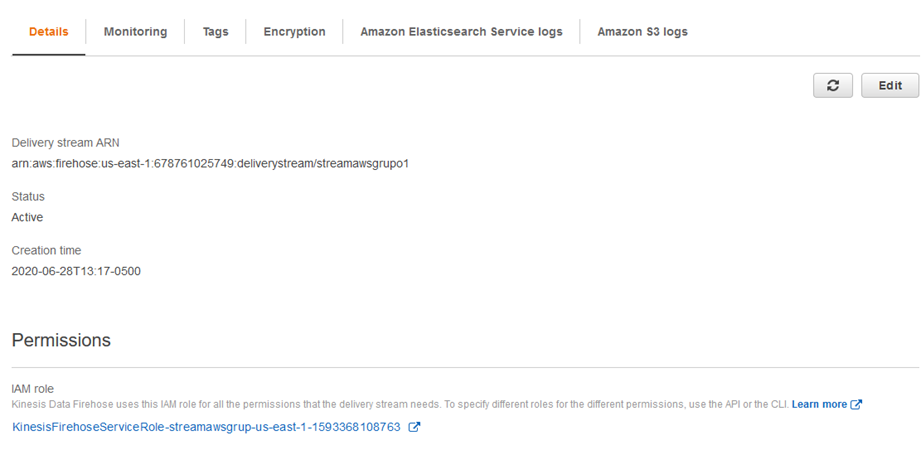


* Paso 4: Se modificó el py, para que se conecte a firehose el cual va a recibir la información en formato json, codificado a 64 bits el mensaje.
* Paso 5: Se ejecuta la función de extracción de tweets.

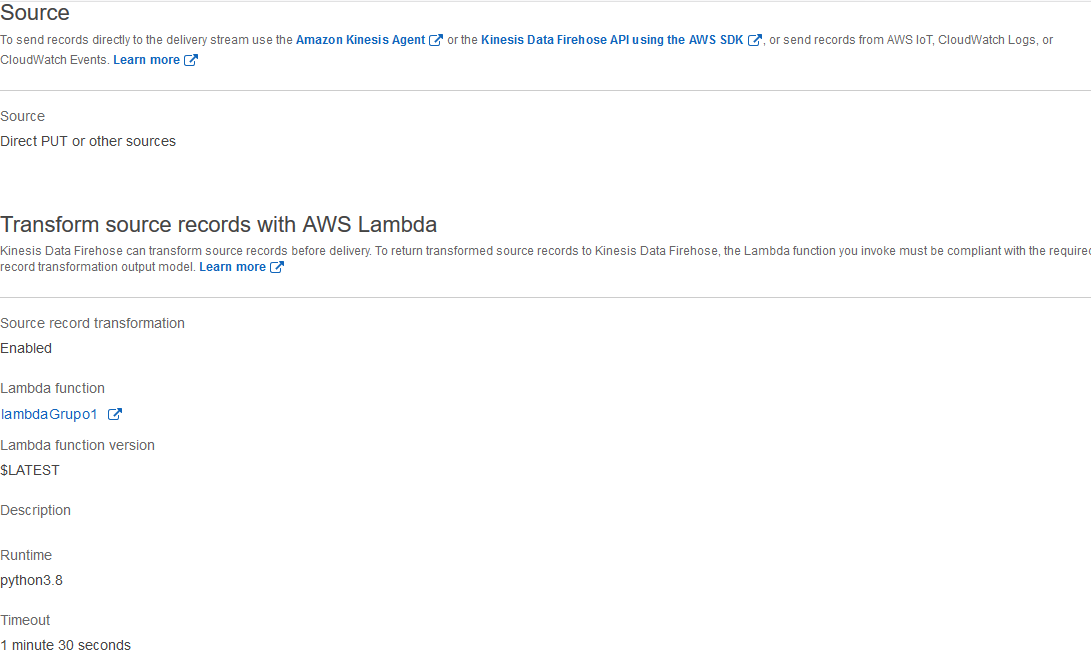


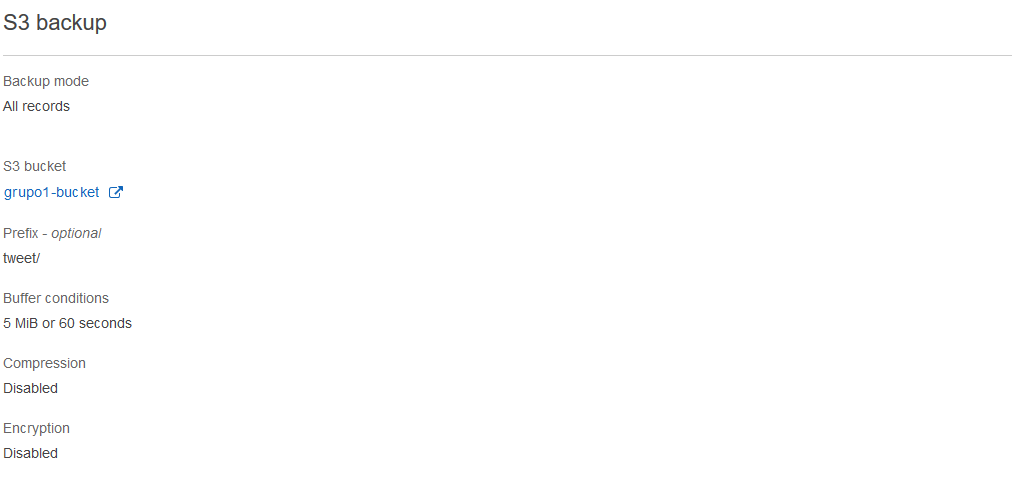
* 1. **Configuración del servicio de Firehose**



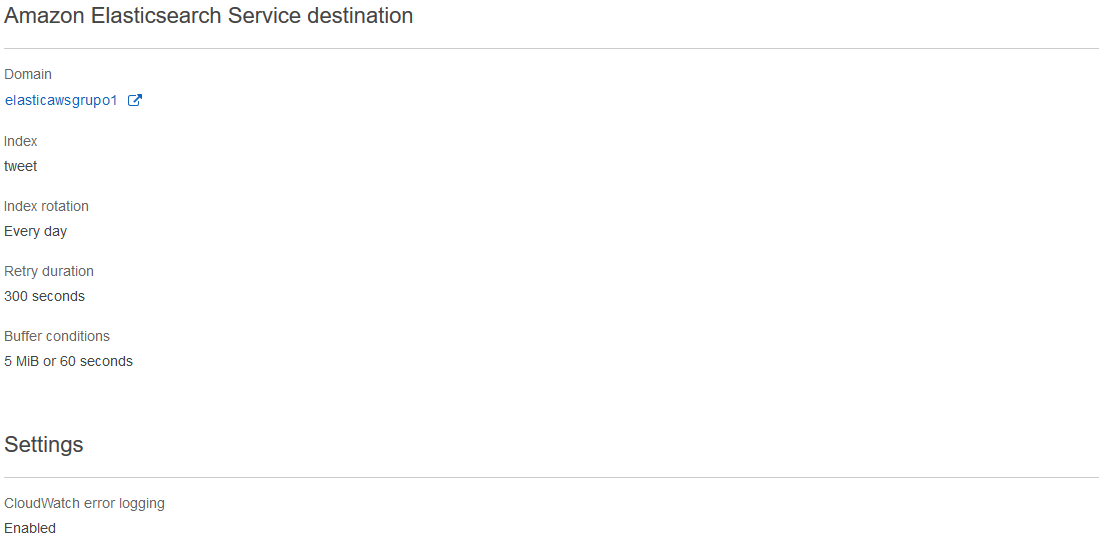


* 1. **Configuración del servicio de Lambda, S3 y Elasticsearch**





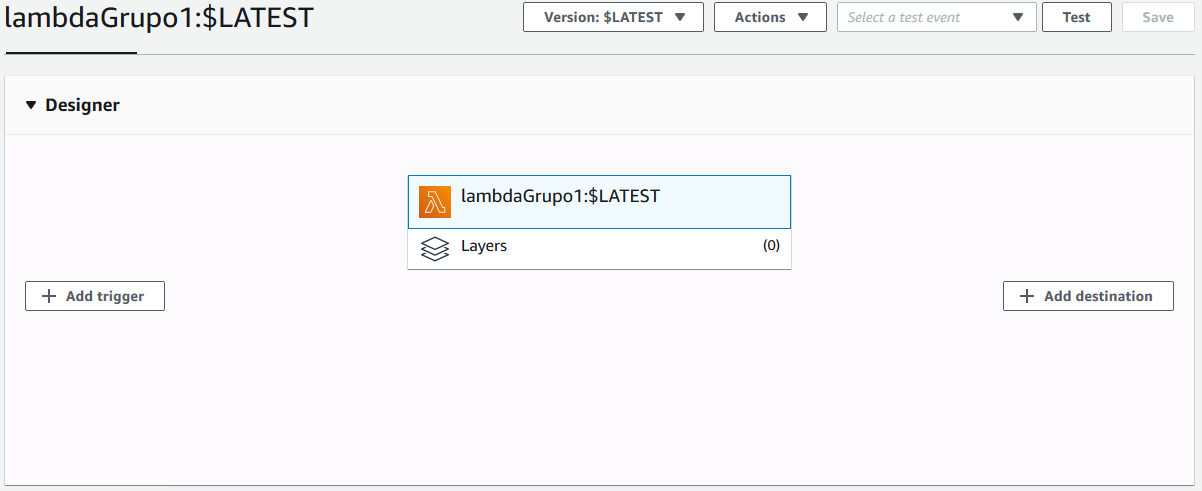
* 1. **Configuración del servicio Elasticsearch**

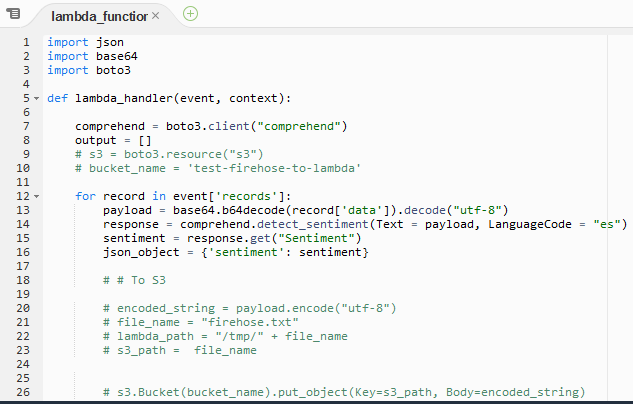


* 1. **Creación de la función Lambda**

Se creó una función lambda que se encarga de utilizar el servicio de comprenhed, para analizar

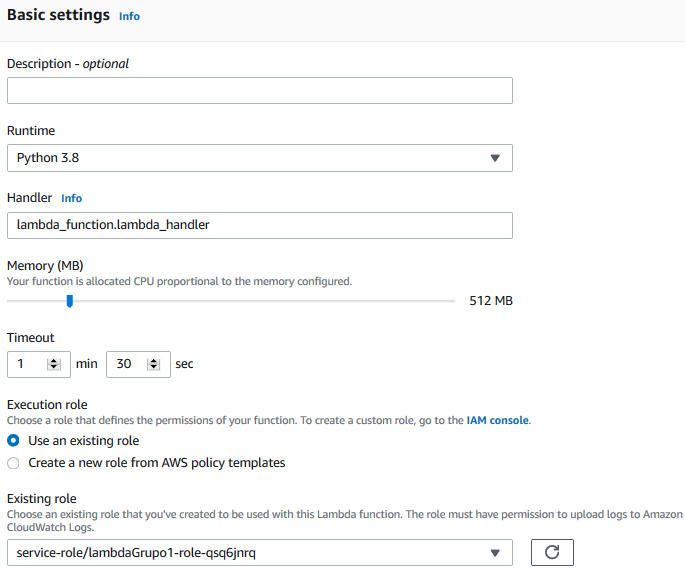
los tweets.



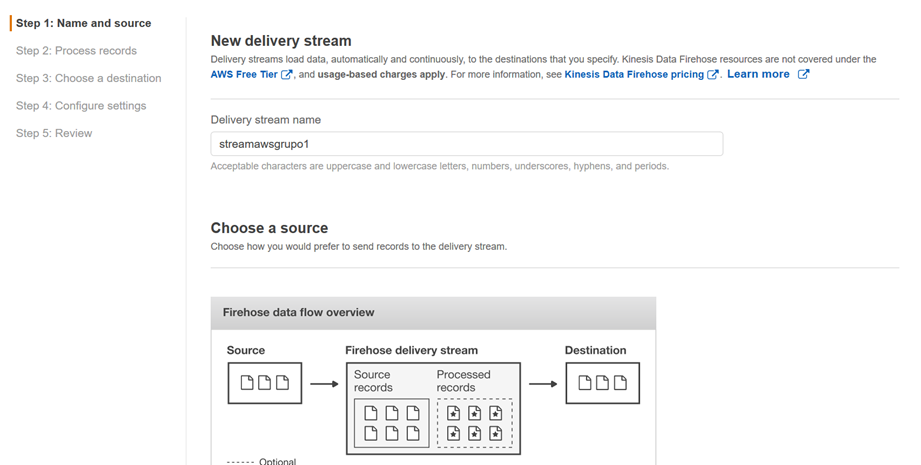


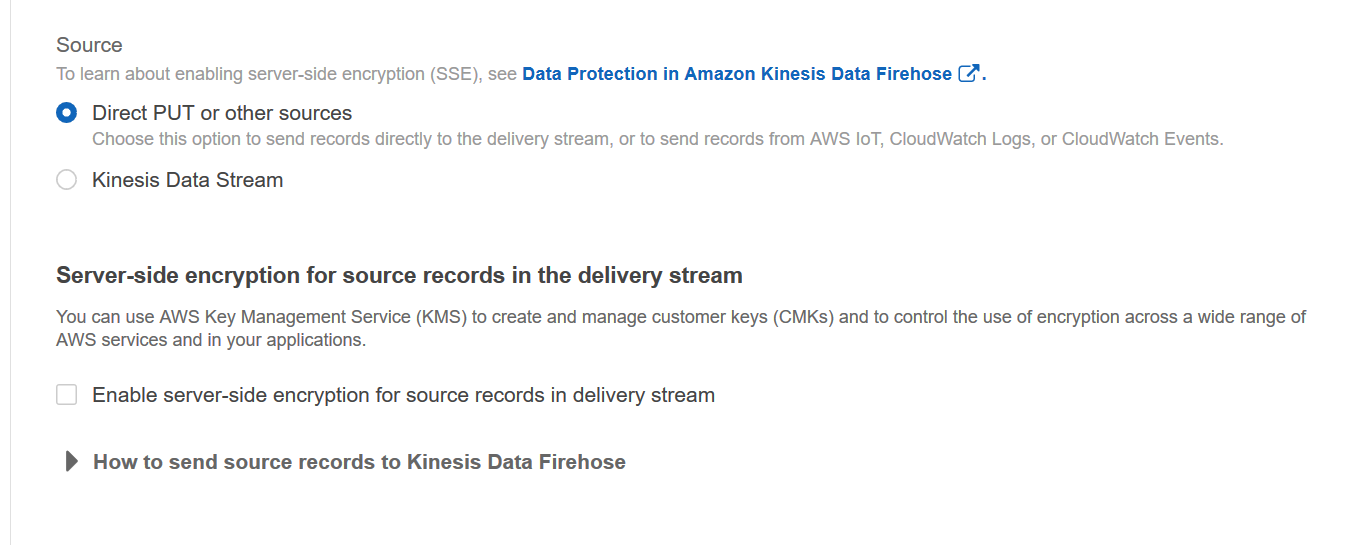
Se modificó la configuración de la función para que durara más tiempo y pueda utilizar más memoria

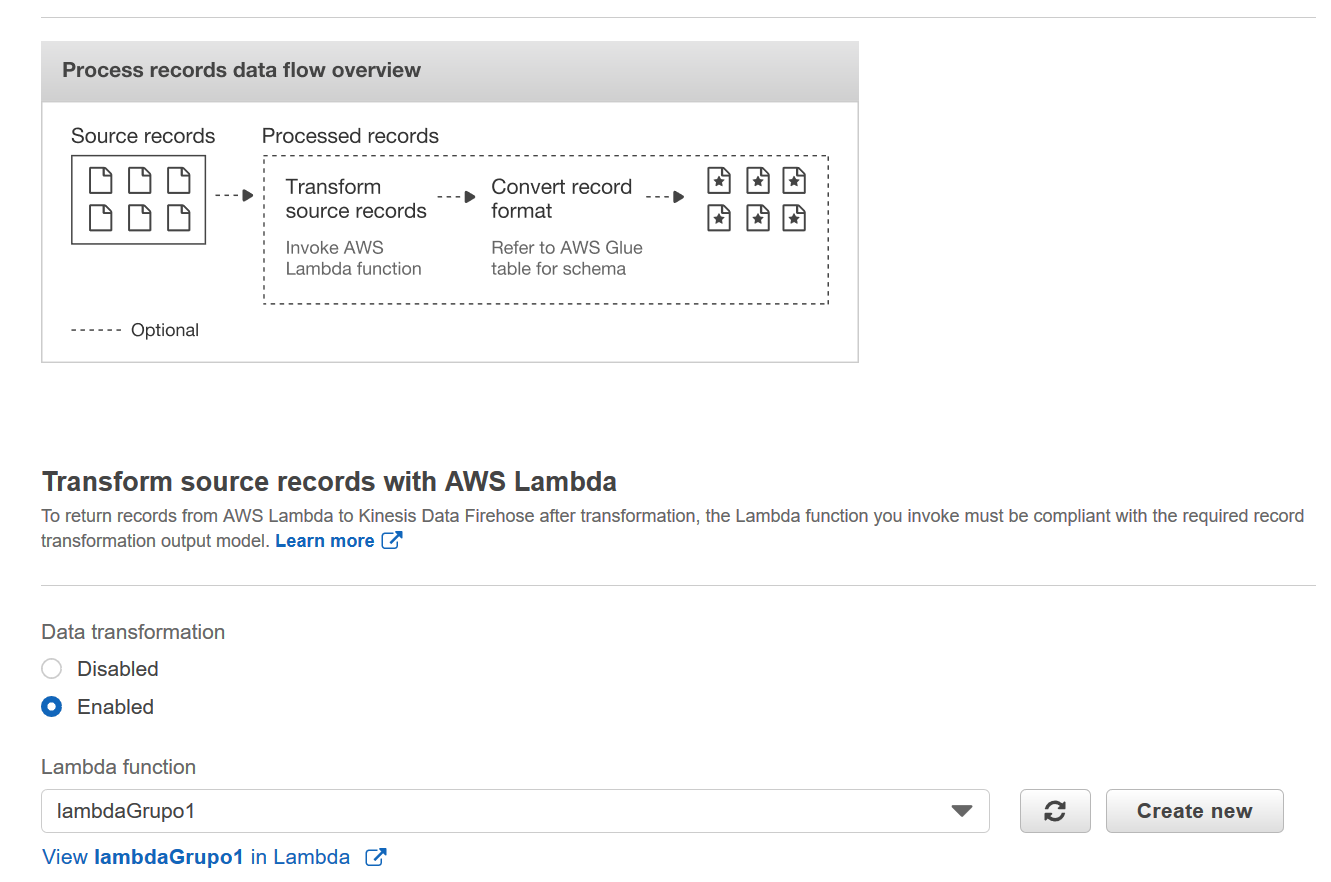
en el procesamiento:

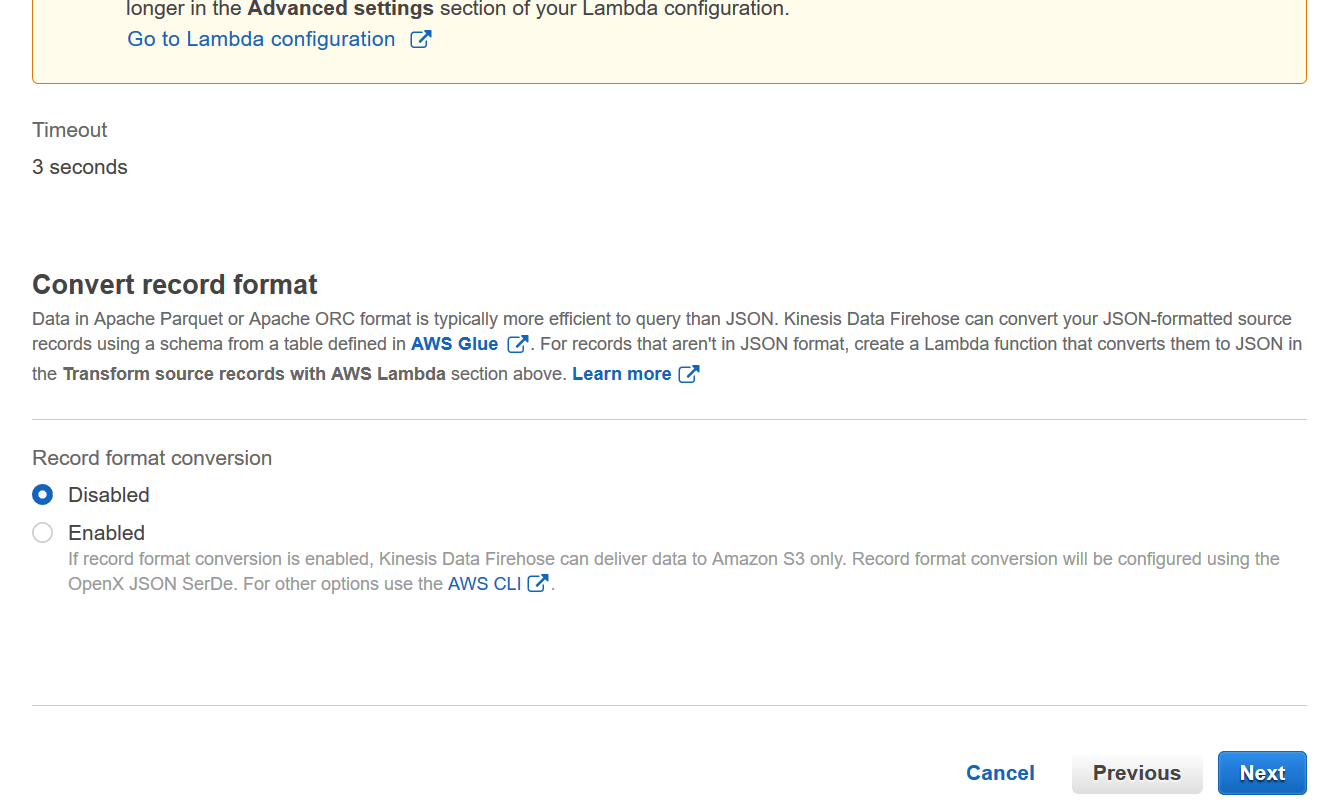


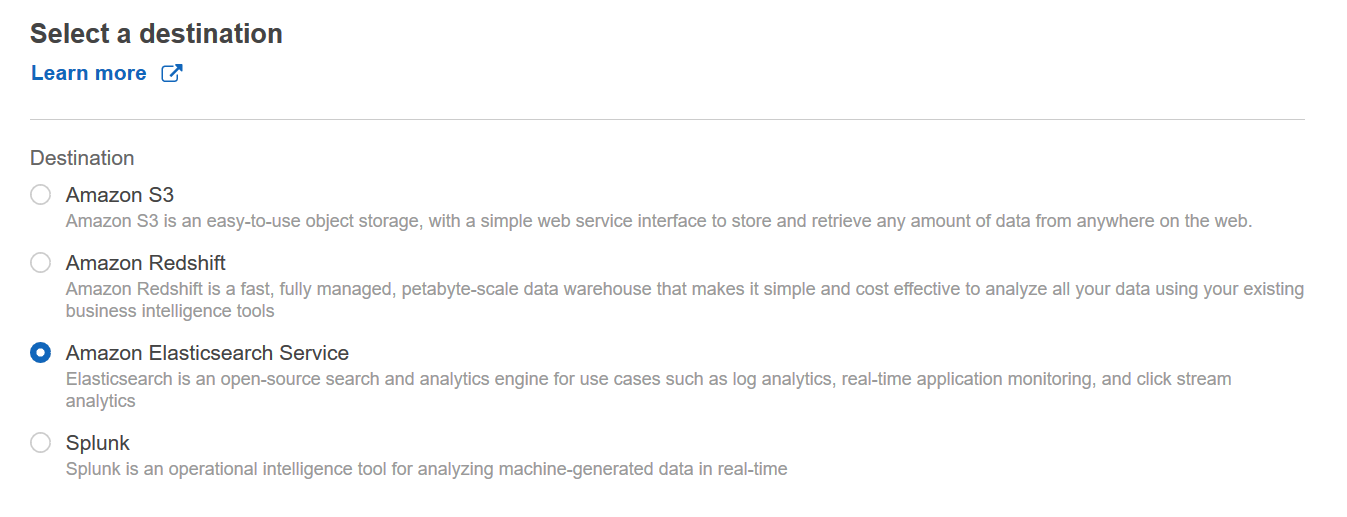
* 1. **Creación del elasticSearch**

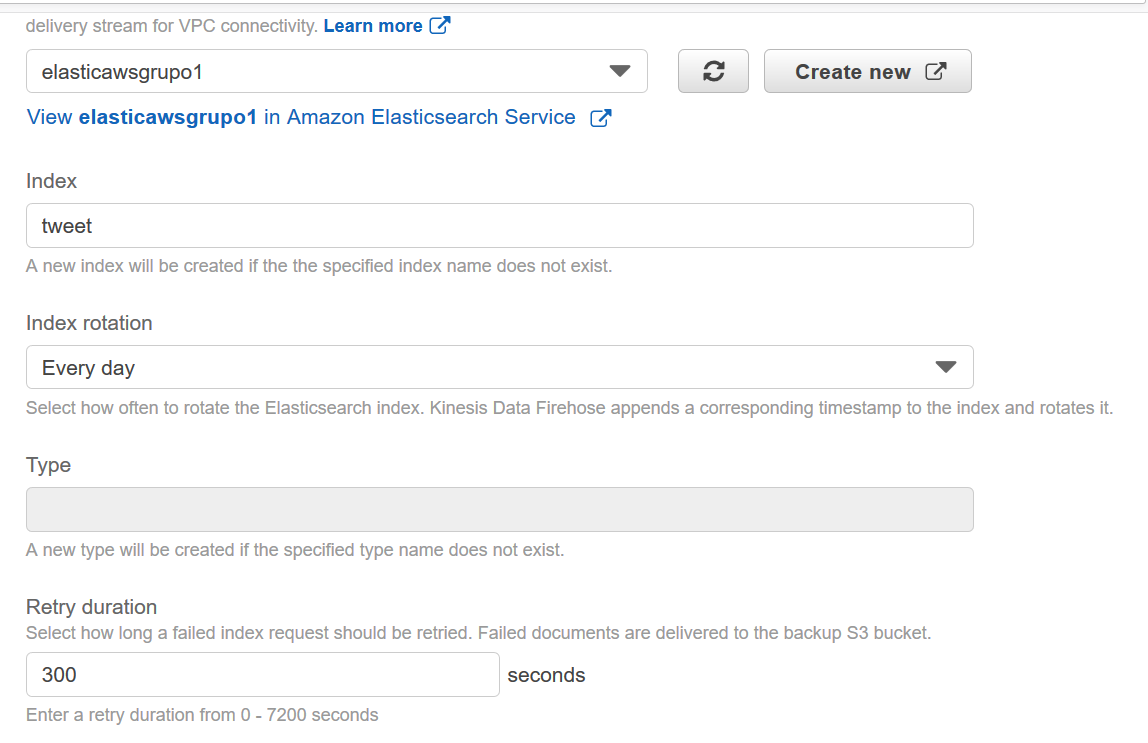
****

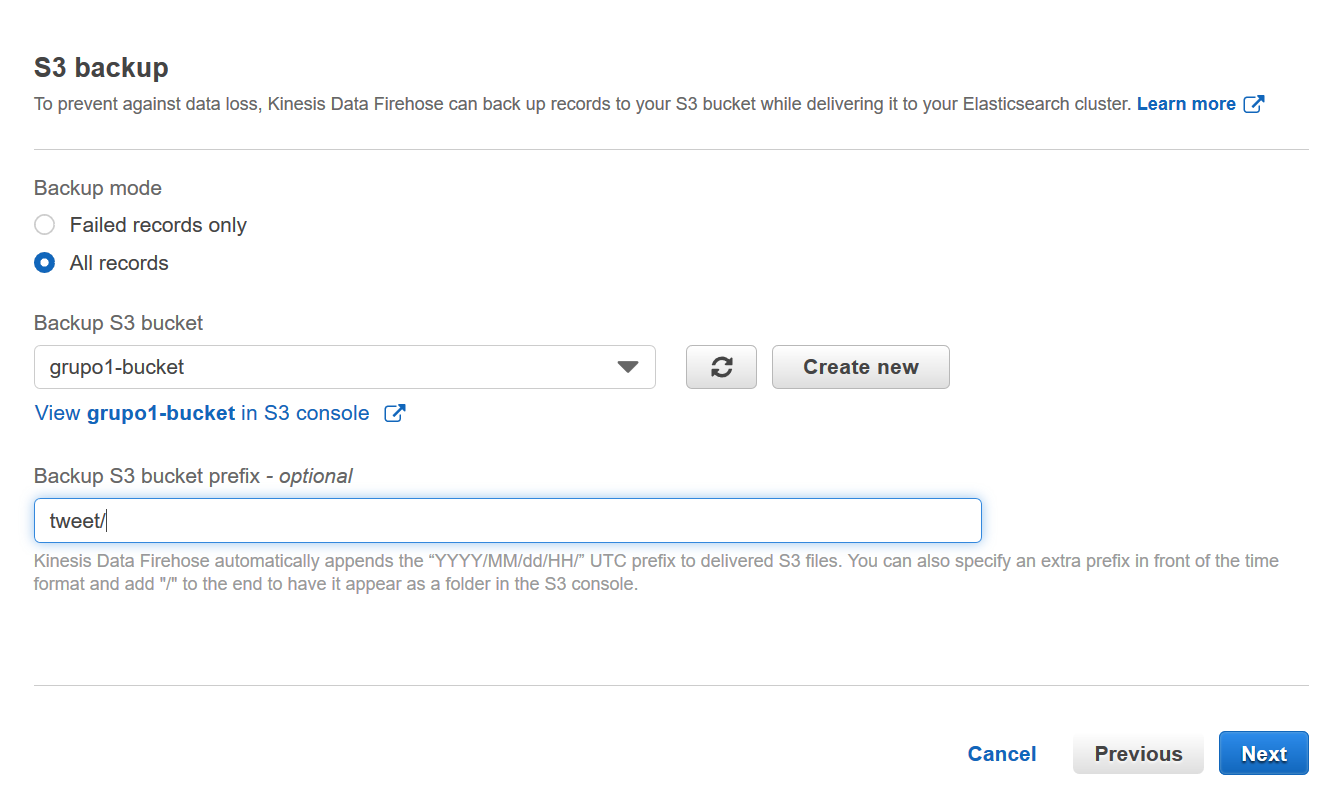


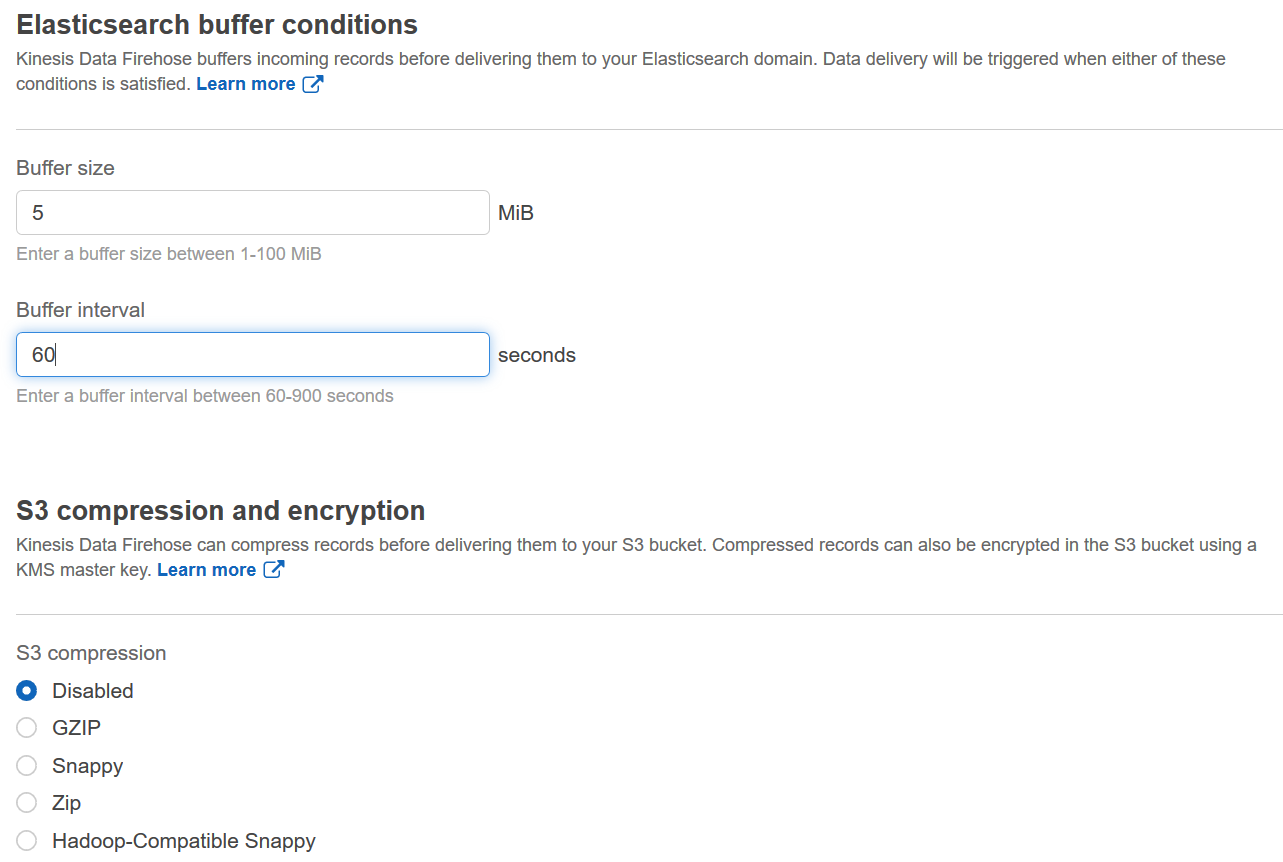


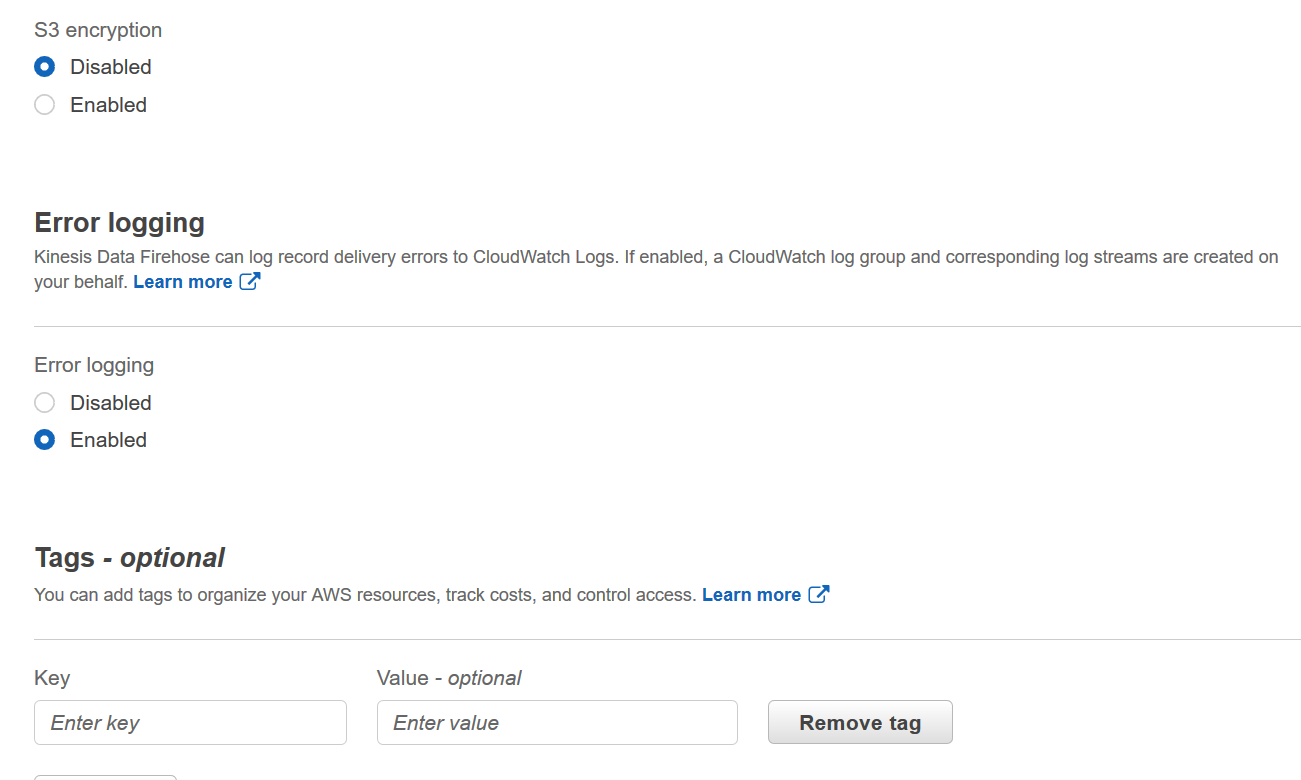


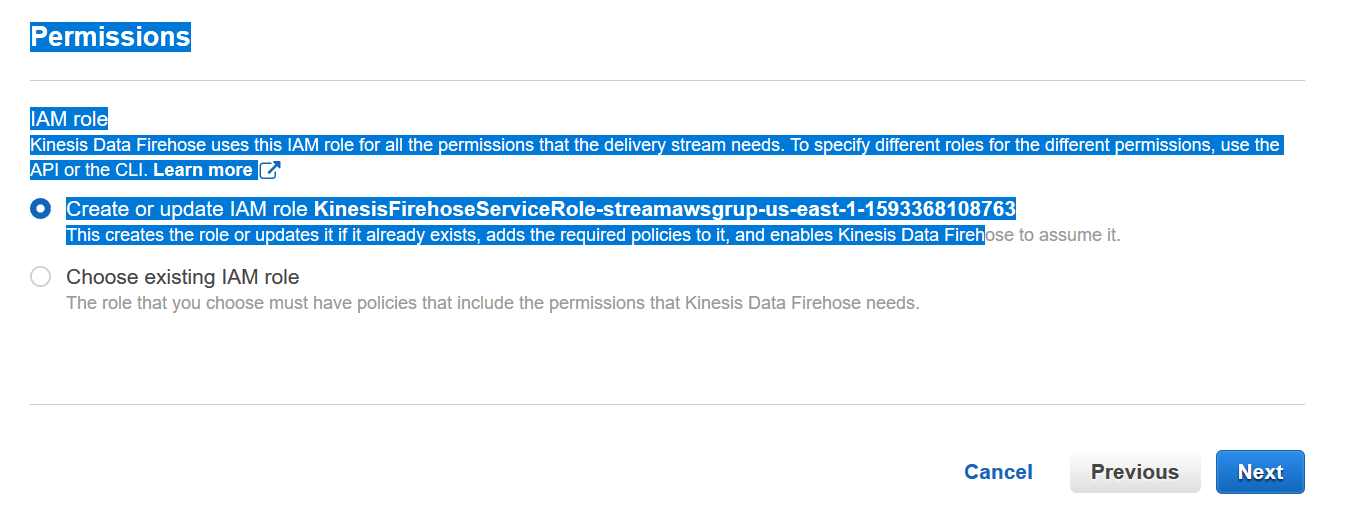


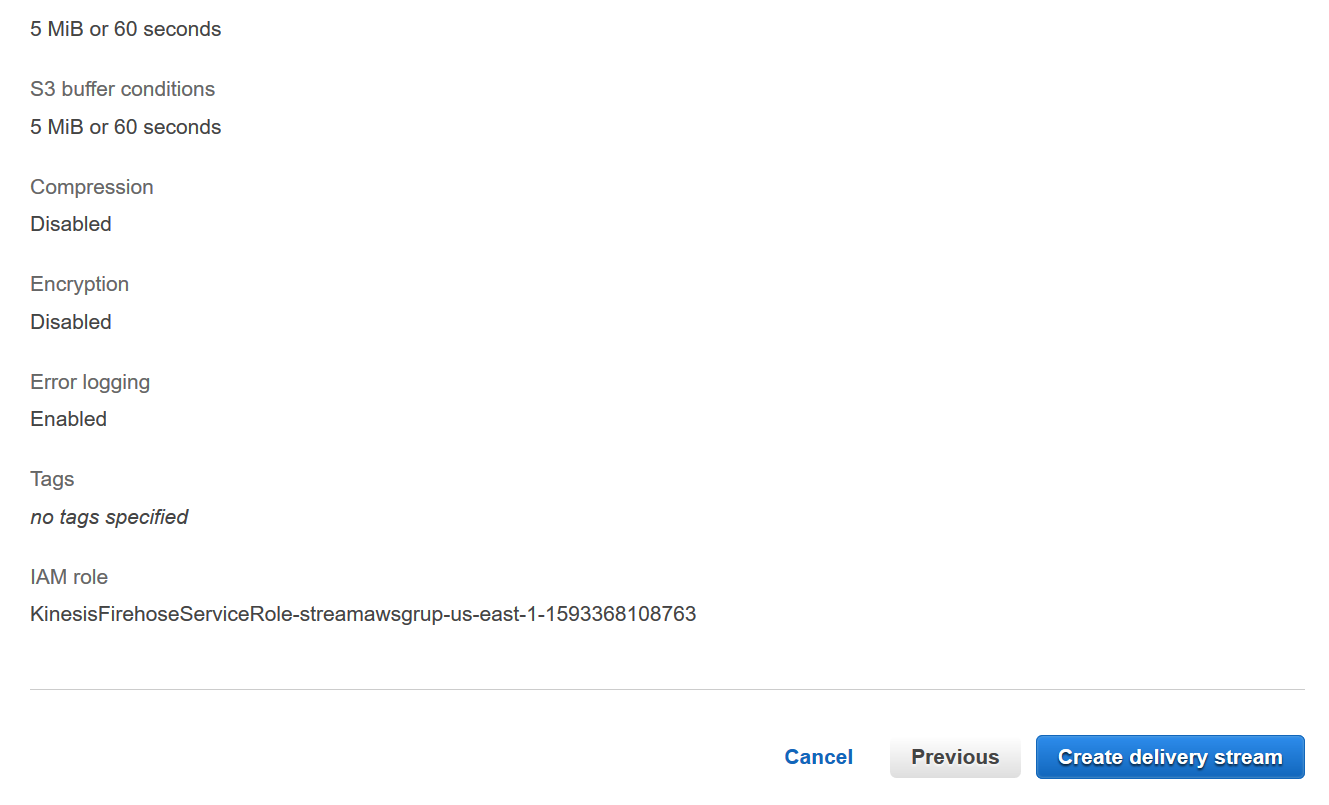




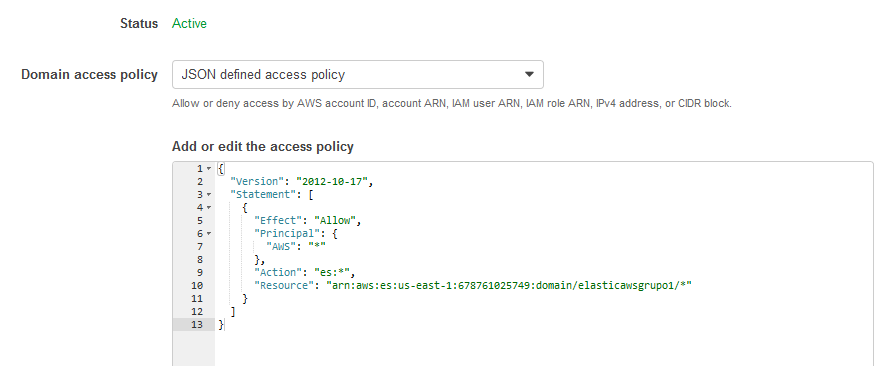








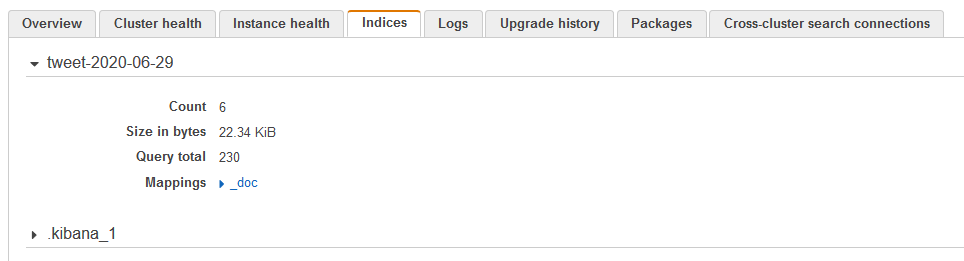
Se modificaron los permisos para acceder al Kibana desde cualquier lugar público



* 1. **Ejecución del proceso de adquisición de tweets**

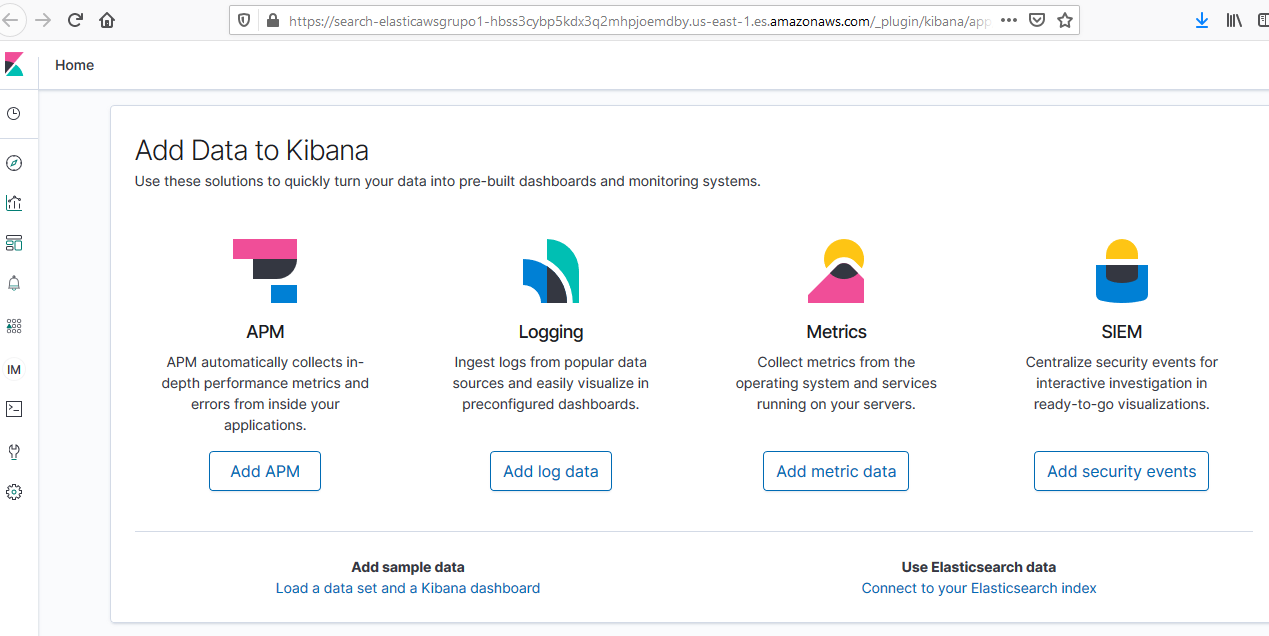


Se valida que se haya creado el índice del elasticsearch

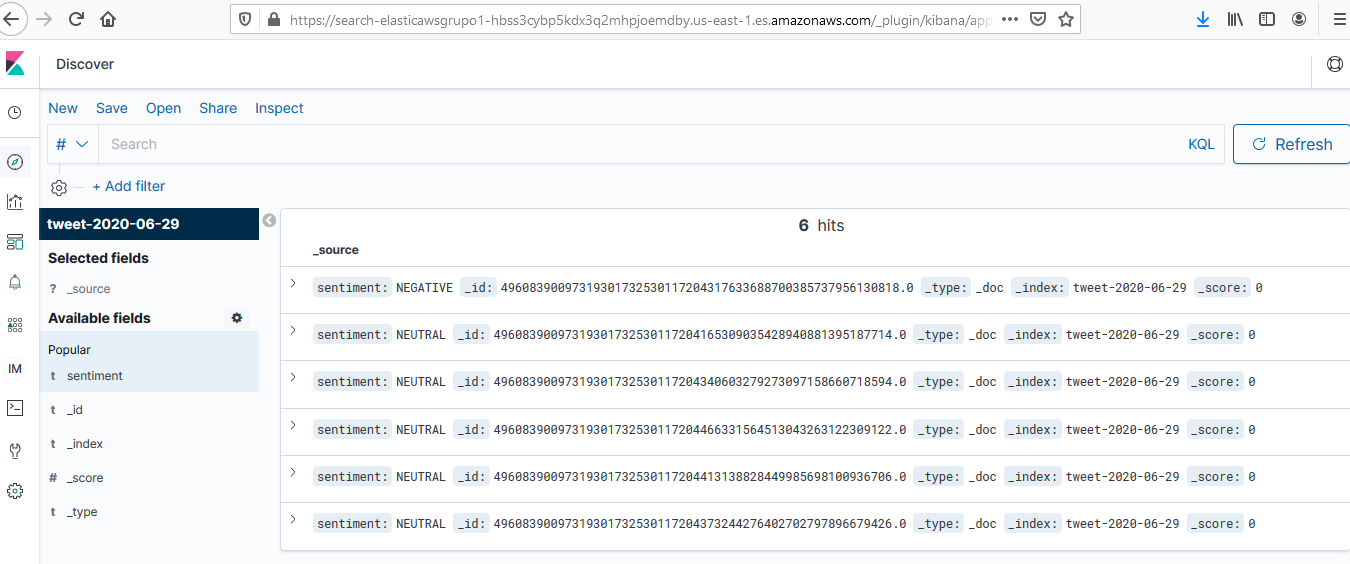


* 1. **Visualización de adquisición de tweets**

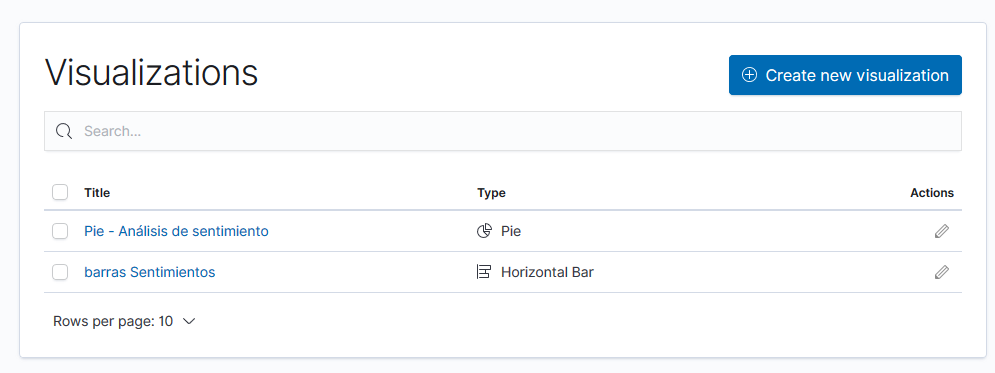
Una vez que se tiene informaciòn en el motor de búsqueda, accedemos a la URL de kibana y podemos configurar el índice de búsquedas.



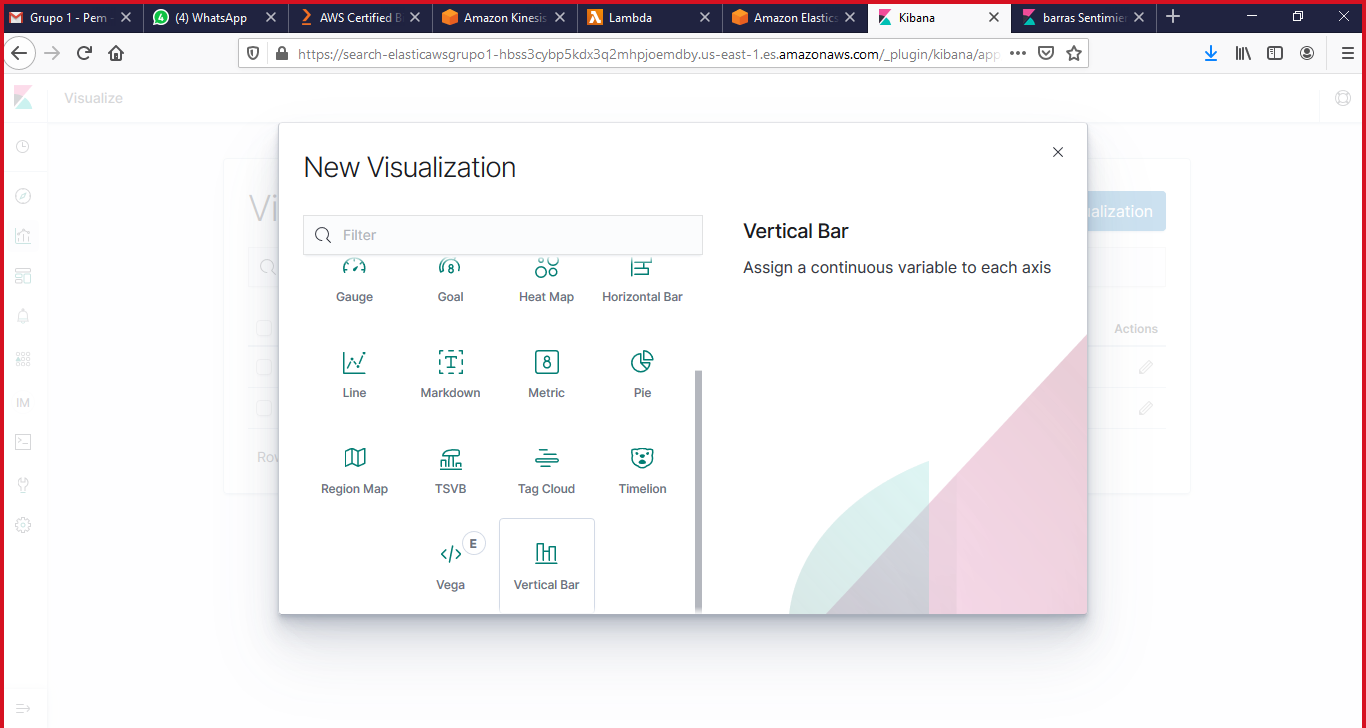
A continuación, en el ícono de discover, podemos validar los datos disponibles para la creación de las gráficas.



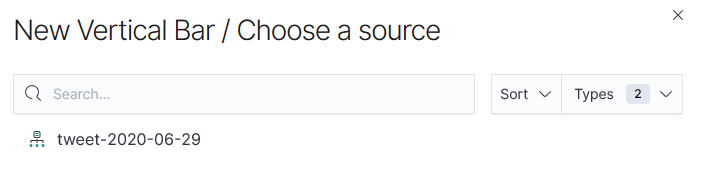
Luego se da clic en el ícono de crear una nueva visualización



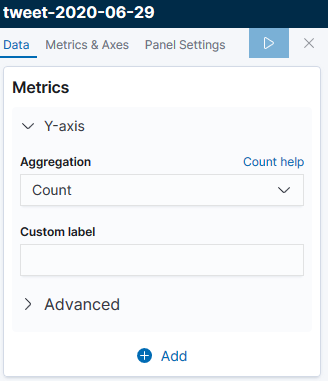
Seleccionamos nueva gráfica de barras



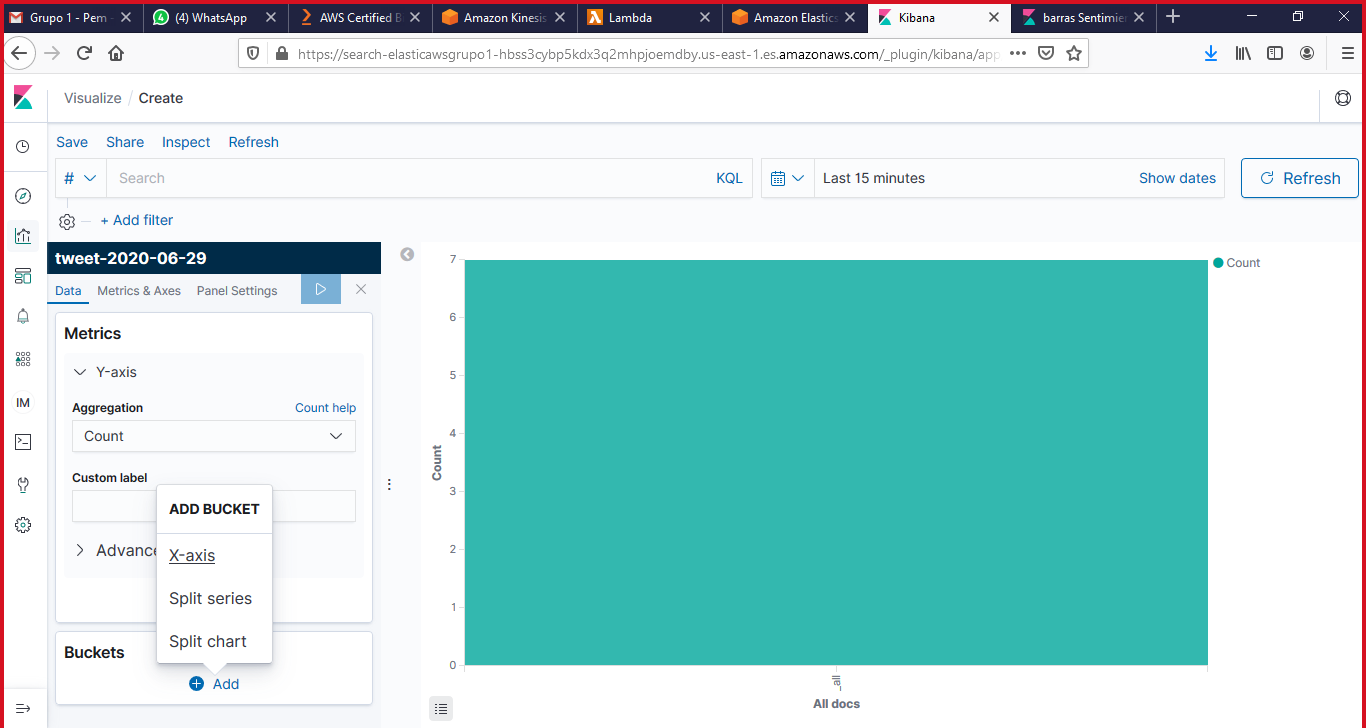
Seleccionamos el set de datos que queremos analizar

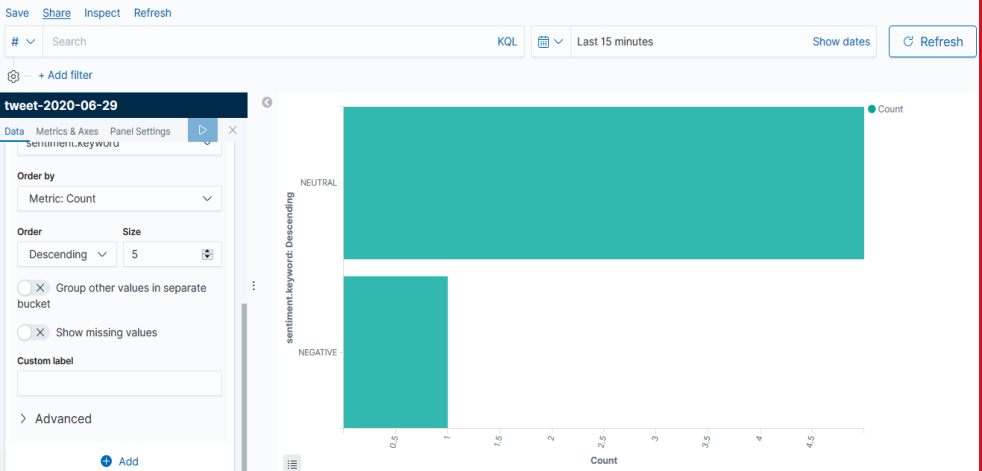
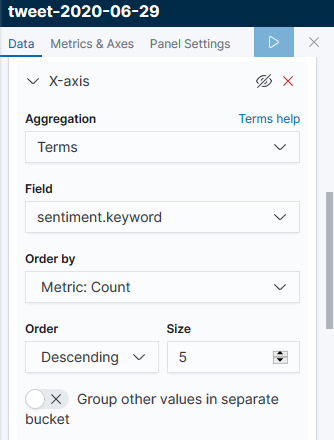


En el eje Y se deja que cuente cuantos registros hay por tipo de respuesta sentimental



A continuación le damos clic en el botón add bucket





De manera similar se pueden replicar los pasos para realizar un gráfico de tipo pie.

