



**UNIVERSIDAD PRIVADA DE TACNA**

**FACULTAD DE INGENIERIA**

**Escuela Profesional de Ingeniería de Sistemas**

**INFORME DE LABORATORIO N° 07  
“INGESTA DE DATOS EN AWS CON KINESIS  
DATA STREAMS”**

Curso: Inteligencia de Negocios

Docente: Ing. Patrick Cuadros Quiroga

**Anahua Huayhua, Jenny Karen (2018062150)**

**Tacna – Perú  
2022**

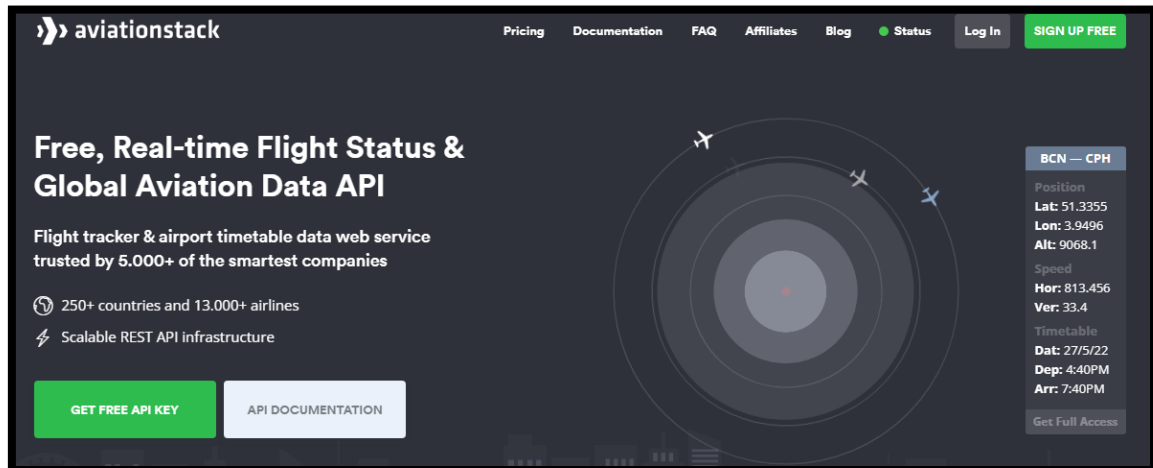


## PRACTICA DE LABORATORIO N° 07

### TEMA: INGESTA DE DATOS EN AWS CON KINESIS DATA STREAMS

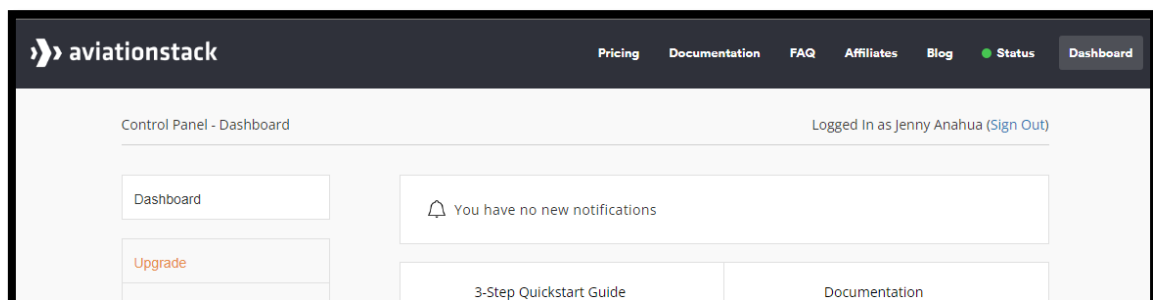
Realizar los siguientes pasos para el laboratorio:

Crearnos una cuenta en <https://aviationstack.com> (3 minutos)



Guardar el token, ya que lo usaremos para obtener datos en tiempo real.

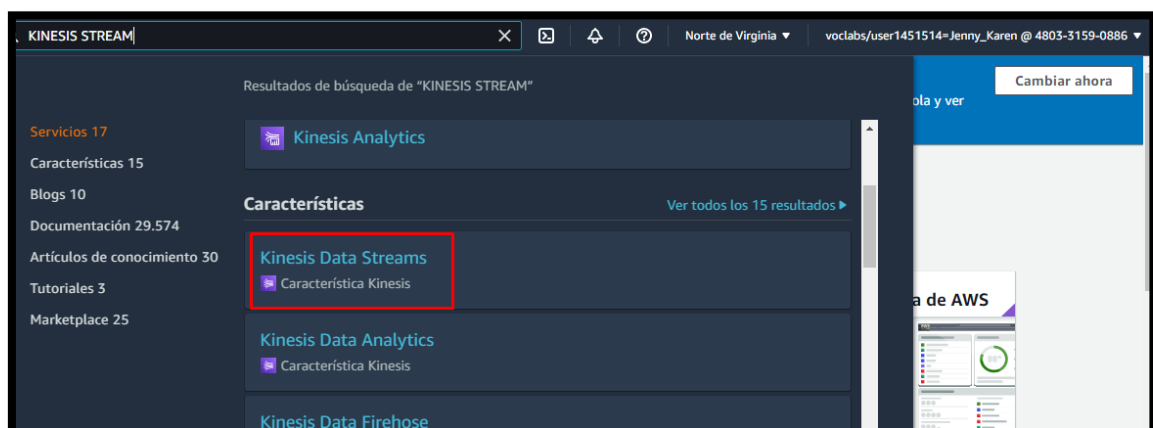
Una vez logueado, entramos a este link, <https://aviationstack.com/dashboard>



Obtendremos los vuelos que se están realizando en tiempo real.

a) Entrar a la consola de AWS.

b) Ir al servicio de Kinesis Data Streams, clic en Crear secuencia de datos





Crear el stream con el nombre de streamVuelos, y luego siguiente y otra vez siguiente. En el campo Number of fragments ingresamos 1 y clic en Crear secuencia de datos.

**Configuración de la secuencia de datos**

Nombre de la secuencia de datos  
streamVuelos  
Los caracteres aceptados son letras mayúsculas y minúsculas, números, guiones bajos, guiones y puntos.

**Capacidad de secuencia de datos** [Info](#) [Solicitar aumento del límite](#)

Los registros de datos se almacenan en Kinesis Data Stream. Una partición es una secuencia de registros de datos identificada de forma exclusiva en una secuencia.

► **Calculador de particiones**

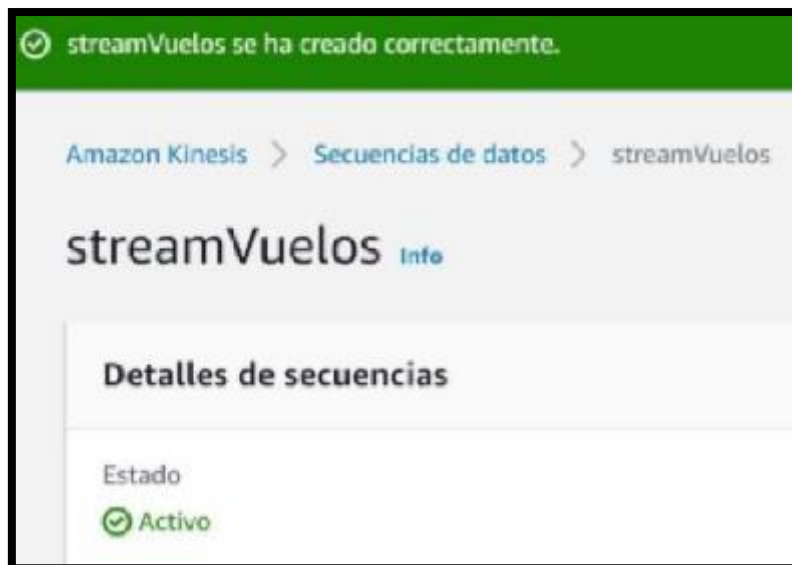
Número de fragmentos abiertos  
Cada partición incorpora hasta 1 MiB/segundo y 1000 registros/segundo, y emite hasta 2 MiB/segundo.  
1  
Mínimo: 1, Máximo: 500, Límite de cuenta: 500.

**Capacidad total de secuencia de datos**  
La capacidad total de secuencia de datos se calcula en función del número de particiones introducidas anteriormente.

Escribir

c) El stream se ha creado

d) Entramos a la EC2 que hemos creado



Ejecutar los siguientes comandos en el terminal del linux:

```
pip3 install pytz --user
```

```
pip3 install requests --uer
```

```
pip3 install wget --user
```



Para ordenar los archivos de los laboratorios, creamos una carpeta llamada 02KDStreamsDescargar los archivos Python:

Subirlos al Cloud9 la siguiente ruta **/home/ec2-user/environment/02KDStreams**

En la línea 12 del archivo WriteFlights.py reemplazarlo por nuestra KEY generada previamente.

Ejecutar el siguiente python:

**python3 WriteFlights.py**

```
ec2-user:~/environment/02KDStreams $ python3 WriteFlights.py
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Ryanair', 'aeropuerto_salida': 'Santander', 'hora_vuelo_salida': '2020-09-01T13:10:00+00:00', 'aeropuerto_llegada': 'El Prat De Llobregat', 'hora_vuelo_llegada': '2020-09-01T14:25:00+00:00'})
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Wizz Air', 'aeropuerto_salida': 'Skavsta', 'hora_vuelo_salida': '2020-09-01T14:40:00+00:00', 'aeropuerto_llegada': 'Liszt Ferenc International', 'hora_vuelo_llegada': '2020-09-01T16:45:00+00:00'})
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Vueling', 'aeropuerto_salida': 'El Prat De Llobregat', 'hora_vuelo_salida': '2020-09-01T15:05:00+00:00', 'aeropuerto_llegada': 'Leonardo Da Vinci (Fiumicino)', 'hora_vuelo_llegada': '2020-09-01T16:55:00+00:00'})
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Vueling', 'aeropuerto_salida': 'El Prat De Llobregat', 'hora_vuelo_salida': '2020-09-01T15:05:00+00:00', 'aeropuerto_llegada': 'Pablo Ruiz Picasso', 'hora_vuelo_llegada': '2020-09-01T16:45:00+00:00'})
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Vueling', 'aeropuerto_salida': 'El Prat De Llobregat', 'hora_vuelo_salida': '2020-09-01T15:05:00+00:00', 'aeropuerto_llegada': 'Stuttgart Echterdingen', 'hora_vuelo_llegada': '2020-09-01T17:10:00+00:00'})
({'fecha_vuelo': '2020-09-01', 'estado_vuelo': 'scheduled', 'aerolinea': 'Ryanair', 'aeropuerto_salida': 'El Prat De Llobregat', 'hora_vuelo_salida': '2020-09-01T15:00+00:00', 'aeropuerto_llegada': 'Nador', 'hora_vuelo_llegada': '2020-09-01T15:50:00+00:00'})
```

Abrir otro terminal, y ejecutar siguiente comando en otra pestaña:

**python3 ReadFlights.py**

```
1T15:10:00+00:00"}
ShardIter : AAAAAAAAAE8IrmP0FEv4SUIFnyZxwPle3abYg3GhbXqxfKG2U/tHh4c0be31TiiU/akir4Igg/QmIN+abM/qL/wTie09ly90Xncx1Gn9rcbt/h3ofV129xx1
z13ZAUjye791ZVQ0Z23aY41Gg9fRfhWgV9eGY1znTG2G9k8Bnzem1RA0xtUfZ33oImfQfngB8H1YcJiHtDphQ+Ur40Vv/DdsVwM
b'{"fecha_vuelo": "2020-09-01", "estado_vuelo": "scheduled", "aerolinea": "KLM", "aeropuerto_salida": "Frankfurt International Airport",
"hora_vuelo_salida": "2020-09-01T14:15:00+00:00", "aeropuerto_llegada": "Schiphol", "hora_vuelo_llegada": "2020-09-01T15:35:00+00:00"
}'
ShardIter : AAAAAAAAAF3fBM9CEu7ESJ/a1fAH4EXaKvq0y1PhyI/FQXQZ1AmTuHK3ae1oIMzurjI6QC2slycGo+XGL+duDvhQpd8yV/Nr7aDFeR+c1N9dzIvoOfu1R18z0B
1FGQVG0MCEAXncbQyRMtCrVmkdrq9LvCPZP8Zcb13pcyMBushsJEqrJ8K1bTwb0T0T1wpaBrXCRzdhwYQjwTEZ/KqeBlepGK+q
b'{"fecha_vuelo": "2020-09-01", "estado_vuelo": "scheduled", "aerolinea": "Lufthansa", "aeropuerto_salida": "Frankfurt International Ai
rport", "hora_vuelo_salida": "2020-09-01T14:10:00+00:00", "aeropuerto_llegada": "John Paul II Balice International", "hora_vuelo_llegad
a": "2020-09-01T15:40:00+00:00"}'
ShardIter : AAAAAAAAAAF1j1j5d6130exjU7BscuA1RWVS1wSK0NVCzR2WOG6pXAd1J1fpuXshJoT4vM40mekXsABQ6xZzzPgtD13ZEDRs0QyeZPhbJlatQCgJaV9jfAR0BI
nBDMtsNtes/rRckfKcZd/bkqDoyJAQcdRBU3o425JyLRFUMLz+M7reinZ1JhC0ptpgAltV/+QCLIV2JbKGuflqn1E01yfrSbks0
b'{"fecha_vuelo": "2020-09-01", "estado_vuelo": "scheduled", "aerolinea": "Cargojet", "aeropuerto_salida": "Vancouver International", "
hora_vuelo_salida": "2020-09-01T06:00:00+00:00", "aeropuerto_llegada": "John C. Munroe", "hora_vuelo_llegada": "2020-09-01T13:30:00+00:
00"}'
```

## Conclusión

Mediante el presente trabajo, se desarrollo los pasos para la creación de un stream, usando el servicio de kinesis data streams.