

viu  
.es

2024 - 2025



# ACTIVIDAD 2

Máster en Big Data y Data Science

02MBID –Sistemas de Almacenamiento y Gestión Big Data

Nombre: Gonzalo Antonio Delgado Rubio

Fecha: 13/05/2024

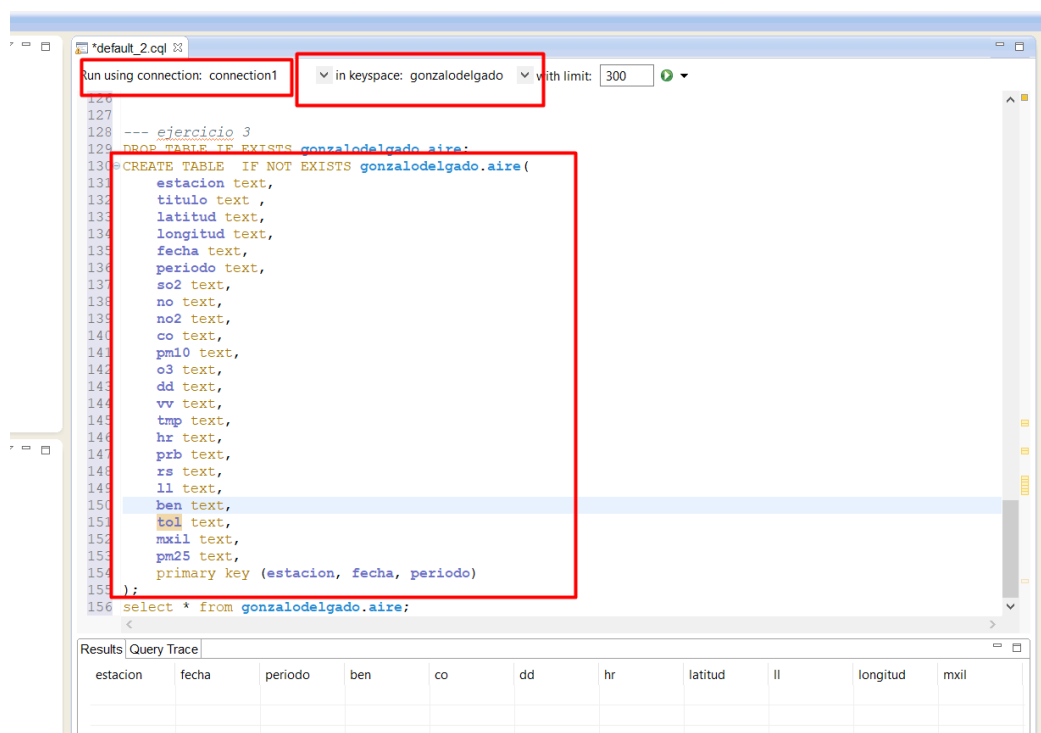
# 1. DSBULK: Migración de Datos

Para esta actividad se empleo la herramienta DSBULK para la carga de información del archivo calidadAire.csv a la tabla en casandra 'aire'. Para ello, se siguieron los siguientes pasos.

## 1. Creación de Tabla Casandra

Se emplea el comando de creación de tabla brindado en la actividad sobre la herramienta devcenter.

En la imagen se muestra cómo se realizó la ejecución de esta.

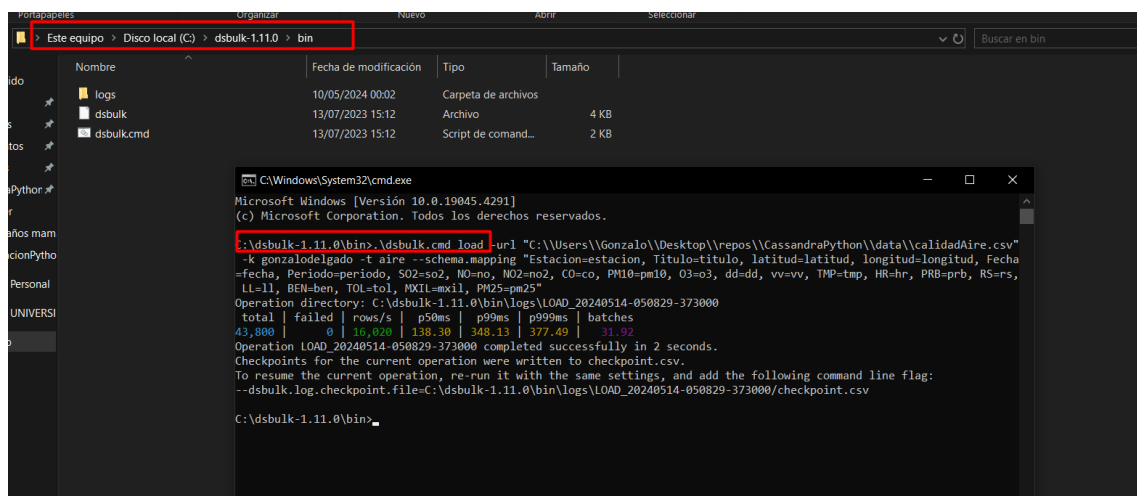


## 2. Comandos de carga dsbulk

Para la carga de nuestro archivo 'calidadAire.csv' a la tabla aire. Empleamos la herramienta dsBulk por lo que emplearemos los siguientes comandos sobre nuestra terminal cmd en Windows. (Pasos similares serían necesarios para Linux o mac)

```
.\dsbulk.cmd load -url
"C:\\Users\\Gonzalo\\Desktop\\repos\\CassandraPython\\data\\calidadAire.csv" -k gonzalodelgado -t aire --schema.mapping "Estacion=estacion,
Titulo=titulo, latitud=latitud, longitud=longitud, Fecha=fecha,
Periodo=periodo, SO2=so2, NO=no, NO2=no2, CO=co, PM10=pm10, O3=o3, dd=dd,
vv=vv, TMP=tmp, HR=hr, PRB=prb, RS=rs, LL=ll, BEN=ben, TOL=tol,
MXIL=mxil, PM25=pm25"
```

Salida de ejecución



Para la ejecución de este comando nos hemos dirigido a la carpeta dsbulk\bin (también podría crearse una variable del sistema) para iniciar la ejecución.

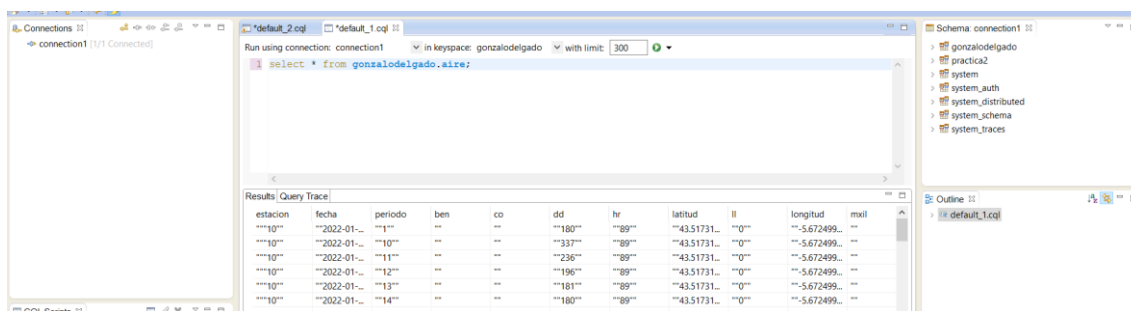
Posterior a ello se ejecutó el comando:

- load -> encargado de la carga
- url <path\_archivo\_a\_cargar> -> establecemos la ruta a archivo a ser cargada
- -k <keyspace\_cassandra> -> colocamos nombre keyspace en cassandra a emplear
- -t <table\_name> -> nombre de la tabla sobre la cual escribiremos
- --schema.mapping "nombre\_columna\_csv: columna\_tabla" -> establecemos las columnas de csv y cassandra a las cuales se deberá escribir los datos.

Finalmente, se nos brinda la confirmación de escritura en tabla

### 3. Comprobación de carga en tabla cassandra

Por último haremos un select sobre cassandra para validar la escritura de los datos



The screenshot shows a database client interface with a query editor and a results pane. The query being executed is:

```
select * from gonzalodelgado.aire;
```

The results pane displays a table with the following columns: `estacion`, `fecha`, `periodo`, `ben`, `co`, `dd`, `hr`, `latitud`, `ll`, `longitud`, and `mml`. The data is presented in a grid format with multiple rows of values.

estacion	fecha	periodo	ben	co	dd	hr	latitud	ll	longitud	mml
10	2022-01-...	1			180	89	43.51731...	0	-5.672499...	
10	2022-01-...	10			337	89	43.51731...	0	-5.672499...	
10	2022-01-...	11			236	89	43.51731...	0	-5.672499...	
10	2022-01-...	12			196	89	43.51731...	0	-5.672499...	
10	2022-01-...	13			181	89	43.51731...	0	-5.672499...	
10	2022-01-...	14			180	89	43.51731...	0	-5.672499...	
10	2022-01-...	15			172	89	43.51731...	0	-5.672499...	