



# **MÁSTER EN BIG DATA Y DATA SCIENCE**

## **02MBID Sistemas de almacenamiento y gestión Big Data**

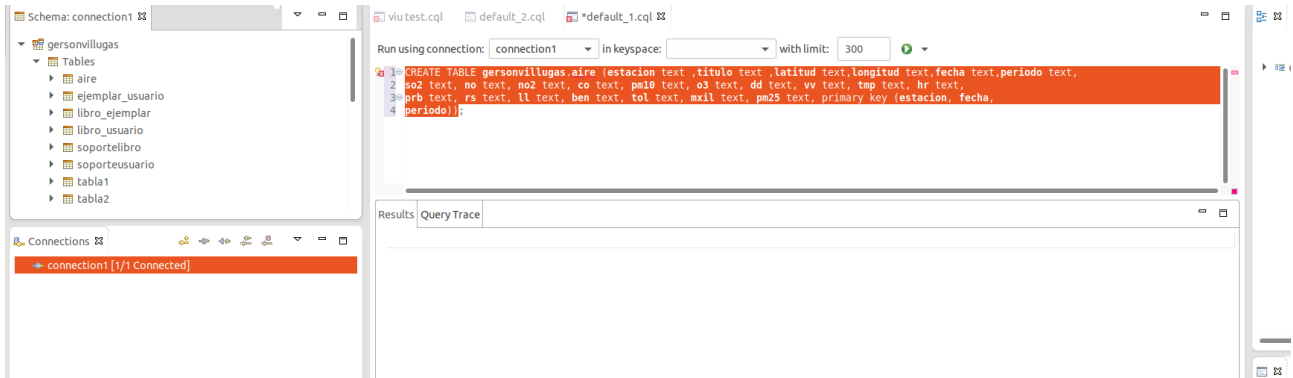
**CURSO 2024-2025**

**ACTIVIDAD 2: migración de datos**

**Hecho por el estudiante Gerson Jordan Villugas Avellaneda**

## paso 1

creación del esquema donde se almacena la información



## instalación del comando

```
(base) ger@ger-B75:~$ export PATH=/home/ger/Documentos/dsbulk-1.11.0/bin:$PATH
(base) ger@ger-B75:~$ source ~/.bashrc
```

## paso 2

Ejecución del comando para carga la data es el siguiente comando

dsbulk load -url /home/ger/Documentos/calidadAire.csv -k gersonvillugas -t aire --schema.mapping "Estacion=estacion,Fecha=fecha,Periodo=periodo,BEN=ben,CO=co,dd=dd,HR=hr,latitud=latitud,LL=ll,longitud=longitud,MXIL=mxil,NO=no,NO2=no2,O3=o3,PM10=pm10,PM25=pm25,PRB=prb,RS=rs,SO2=so2,Titulo=titulo,TMP=tmp,TOL=tol,vv=vv"

```
--dsbulk.log.checkpoint.file=/home/ger/logs/LOAD_20240507-212155-053461/checkpoint.csv
(base) ger@ger-B75:~$ dsbulk load -url /home/ger/Documentos/calidadAire.csv -k gersonvillugas -t aire --schema.mapping "Estacion=estacion,Fecha=fecha,Periodo=periodo,BEN=ben,CO=co,dd=dd,HR=hr,latitud=latitud,LL=ll,longitud=longitud,MXIL=mxil,NO=no,NO2=no2,O3=o3,PM10=pm10,PM25=pm25,PRB=prb,RS=rs,SO2=so2,Titulo=titulo,TMP=tmp,TOL=tol,vv=vv"
Operation directory: /home/ger/logs/LOAD_20240507-212155-053461
total | failed | rows/s | p50ms | p99ms | p999ms | batches
43,806 | 0 | 9.417 | 73.55 | 620.76 | 881.11 | 31.91
Operation LOAD_20240507-212155-053461 completed successfully in 4 seconds.
Checkpoints for the current operation were written to checkpoint.csv.
To resume the current operation, re-run it with the same settings, and add the following command line flag:
--dsbulk.log.checkpoint.file=/home/ger/logs/LOAD_20240507-212155-053461/checkpoint.csv
(base) ger@ger-B75:~$
```

paso 3

comprobar que se cargo la información

5  
6 select \* from gersonvillugas.aire;  
7

ResultsQuery Trace

estacion	fecha	periodo	ben	co	dd	hr	latitud	ll	longitud	mxil	no	nc
""10""	""2022-01-01""	""1""	""	""	""180""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""10""	""	""	""337""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""11""	""	""	""236""	""89""	""43.517315""	""0""	""-5.672499""	""	""4""	""
""10""	""2022-01-01""	""12""	""	""	""196""	""89""	""43.517315""	""0""	""-5.672499""	""	""4""	""
""10""	""2022-01-01""	""13""	""	""	""181""	""89""	""43.517315""	""0""	""-5.672499""	""	""18""	""
""10""	""2022-01-01""	""14""	""	""	""180""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""15""	""	""	""176""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""16""	""	""	""200""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""17""	""	""	""188""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""18""	""	""	""196""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""
""10""	""2022-01-01""	""19""	""	""	""196""	""89""	""43.517315""	""0""	""-5.672499""	""	""1""	""

comprobar la cantidad de los datos ,como vemos se cargaron todo los datos

```
4  -- gersonvillugas.aire;
5
6  select count(*) from gersonvillugas.aire;
7
```

Results Query Trace

count
43.800