

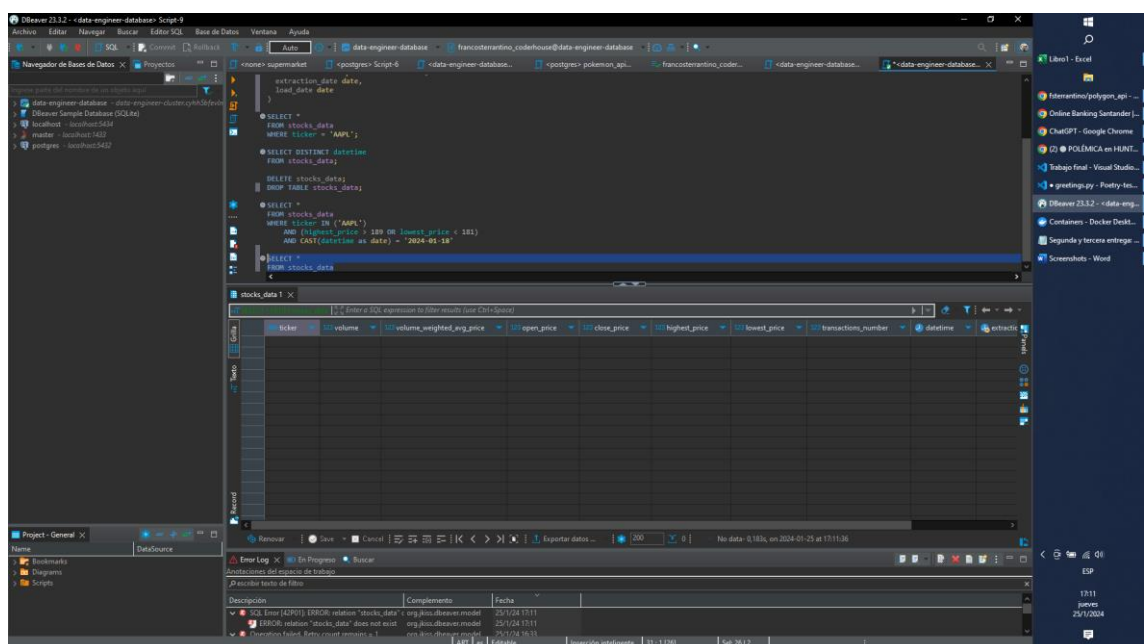
Starting configs

- Config.ini:

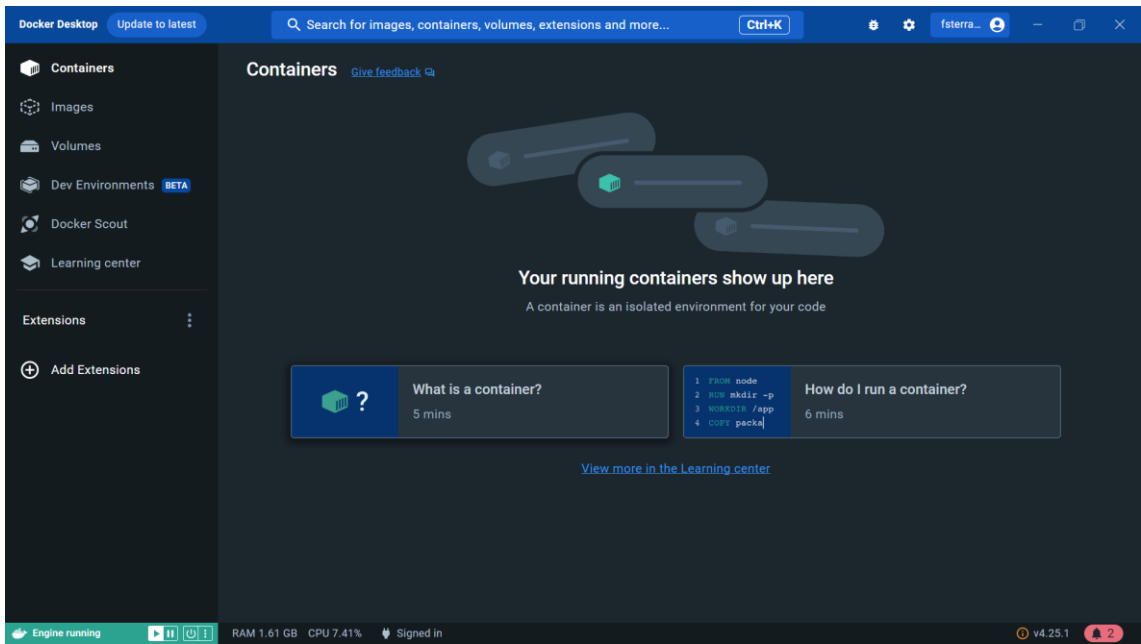
```
[API_PARAMETERS]
tickers_to_query_list = AAPL, AMZN, TSLA, BABA, MSFT, ACN, T, BAC, KO
time_frame = hour
frame_multiplier = 1
```

This config will query for that tickers info in an hourly bar.

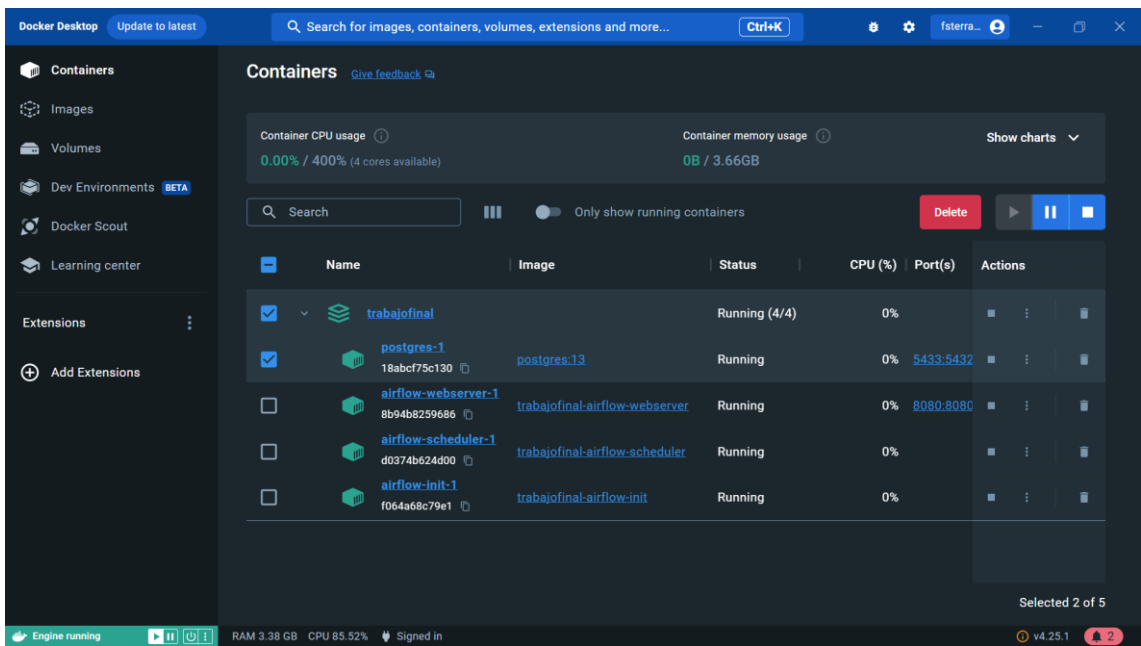
- Checking empty DB table:



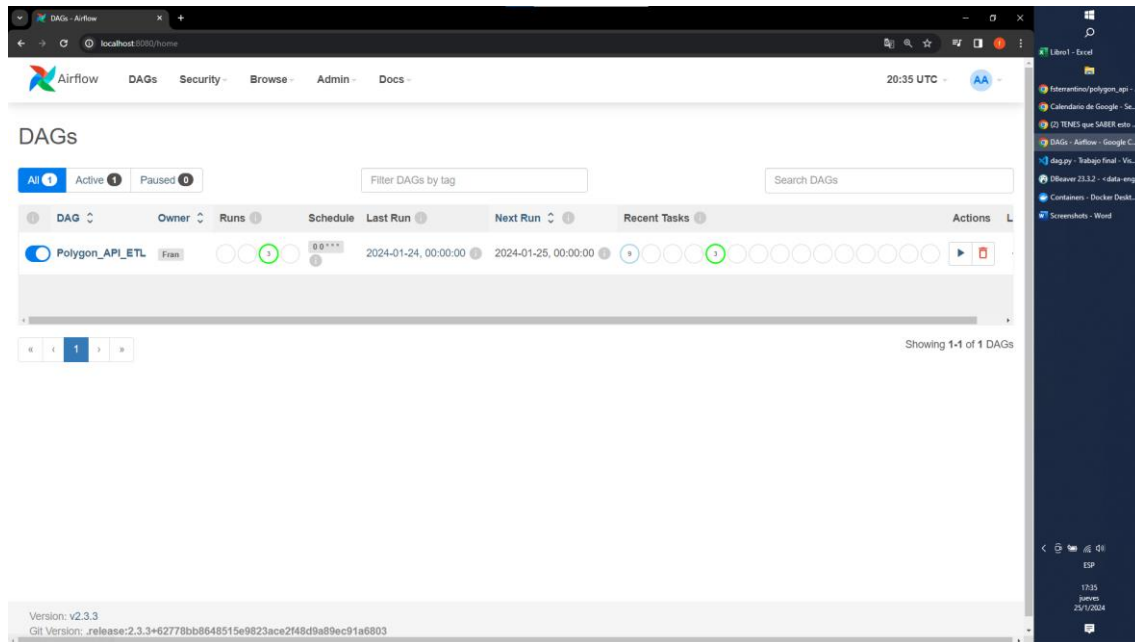
- Checking no containers in Docker:



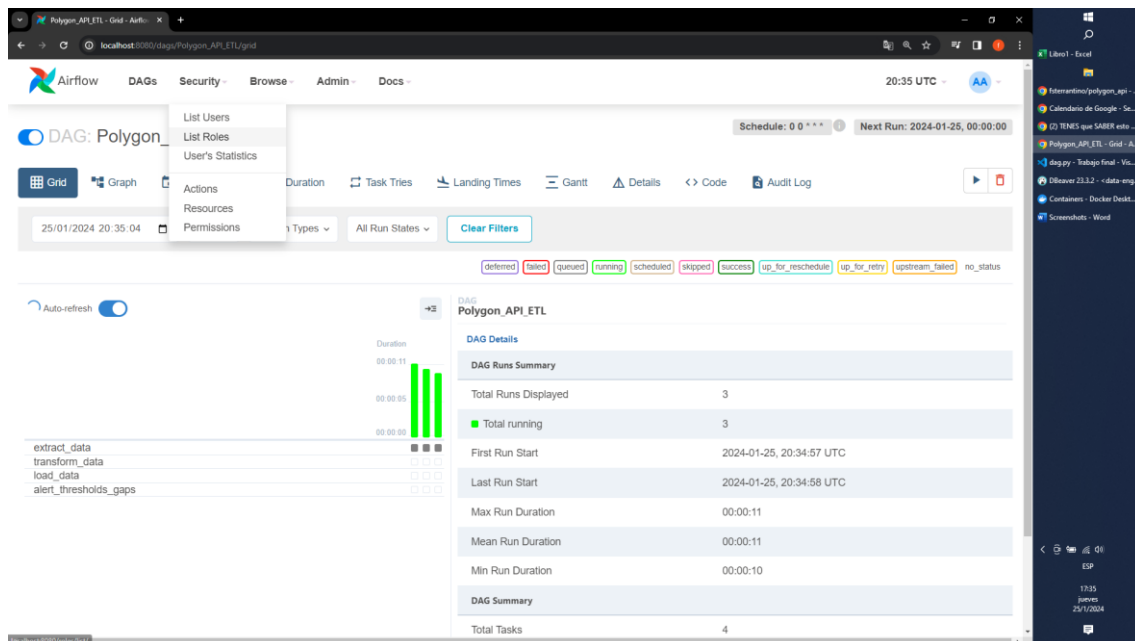
- Container started



Airflow

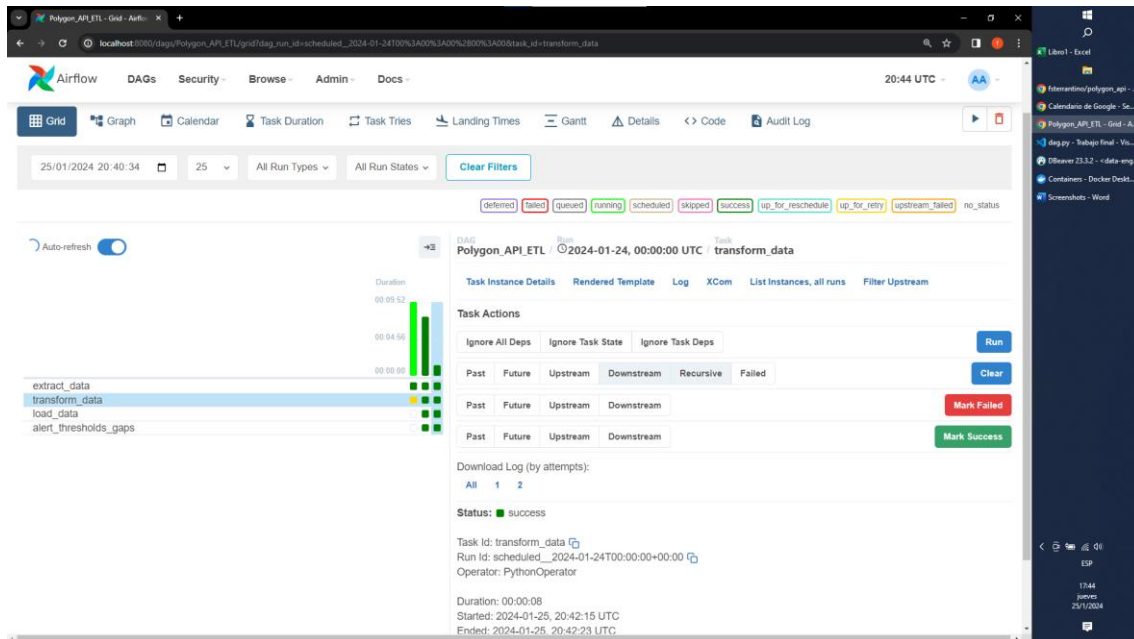


- Starting DAG execution (from startdate: 22/01/2024 to 24/01/2024)



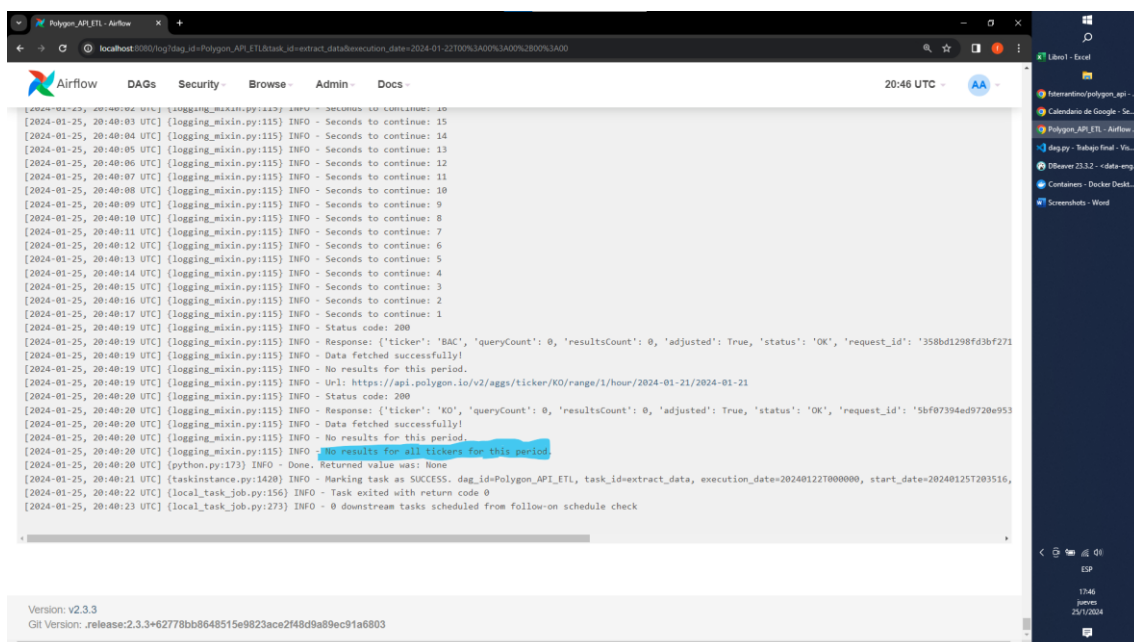
The expected result is the first transform to fail because it's retrieving data from Sunday, when stocks markets are closed.

The extraction delays some time because the API has a rate-limit of 5 requests per minute.

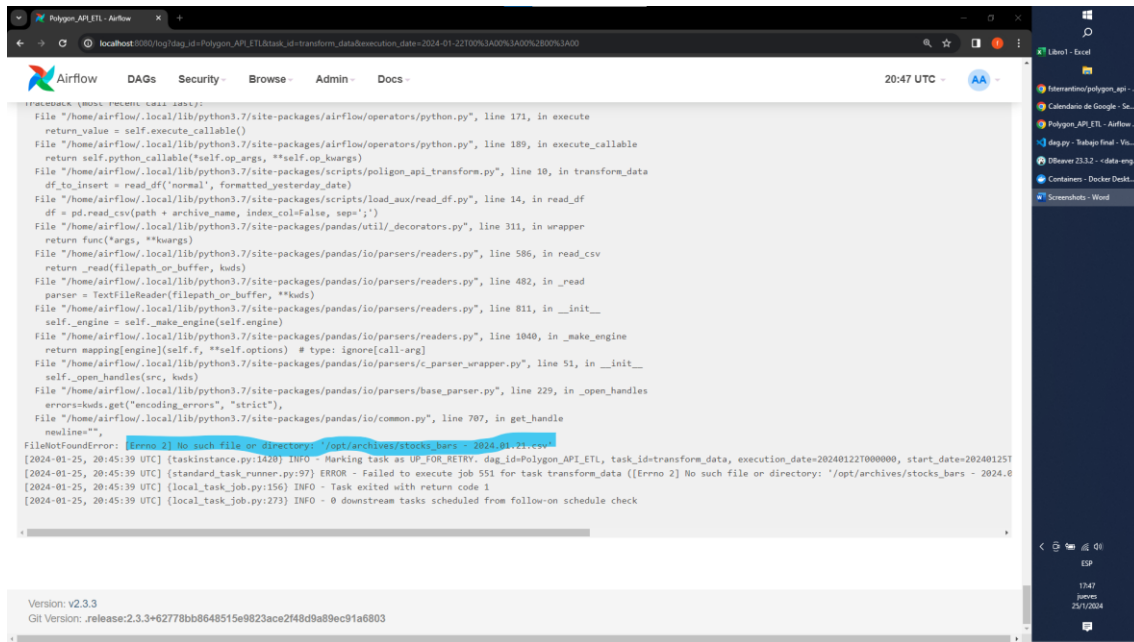


22/01/2024: Failed in transformation task as expected. The extraction did not generate the file as there is no activity on Sunday.

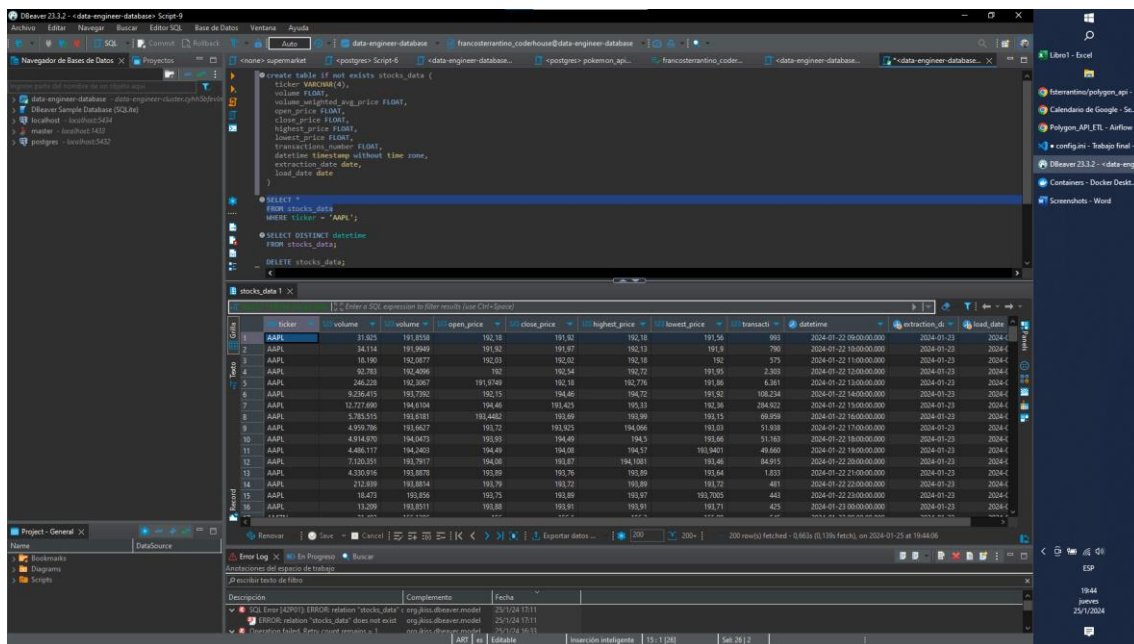
Extraction logs:



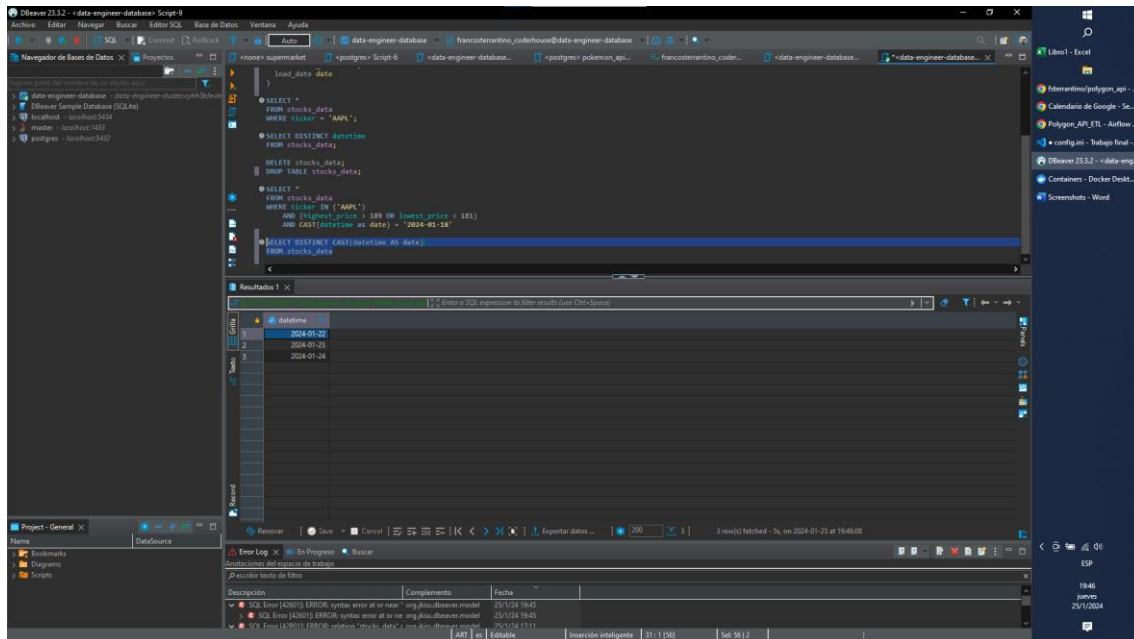
Transform logs:



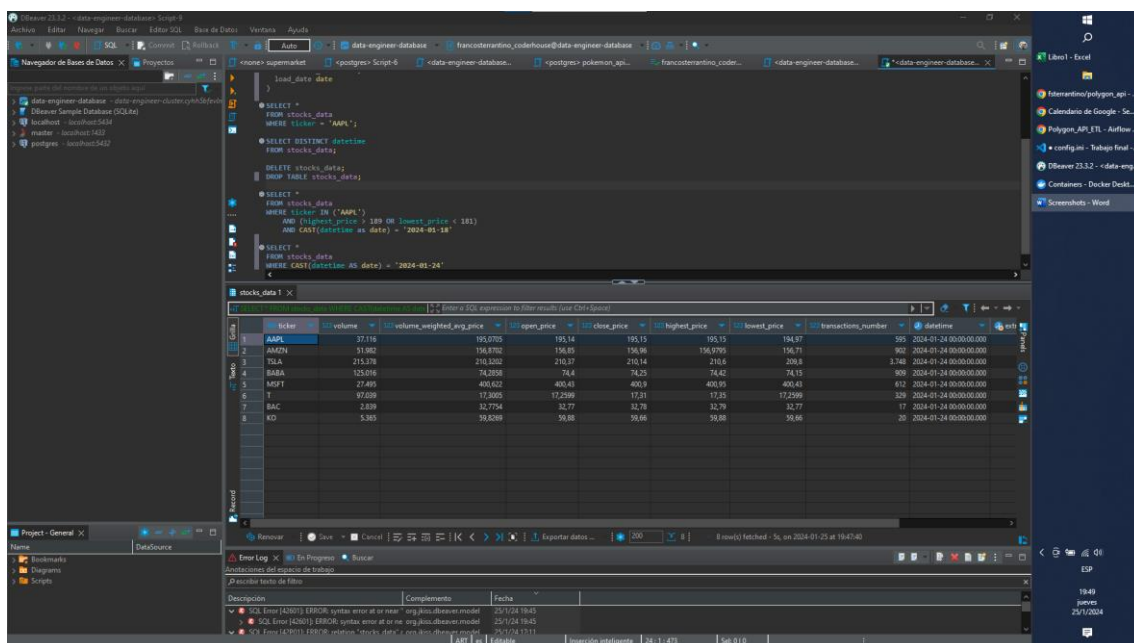
Load impact in DDBB:



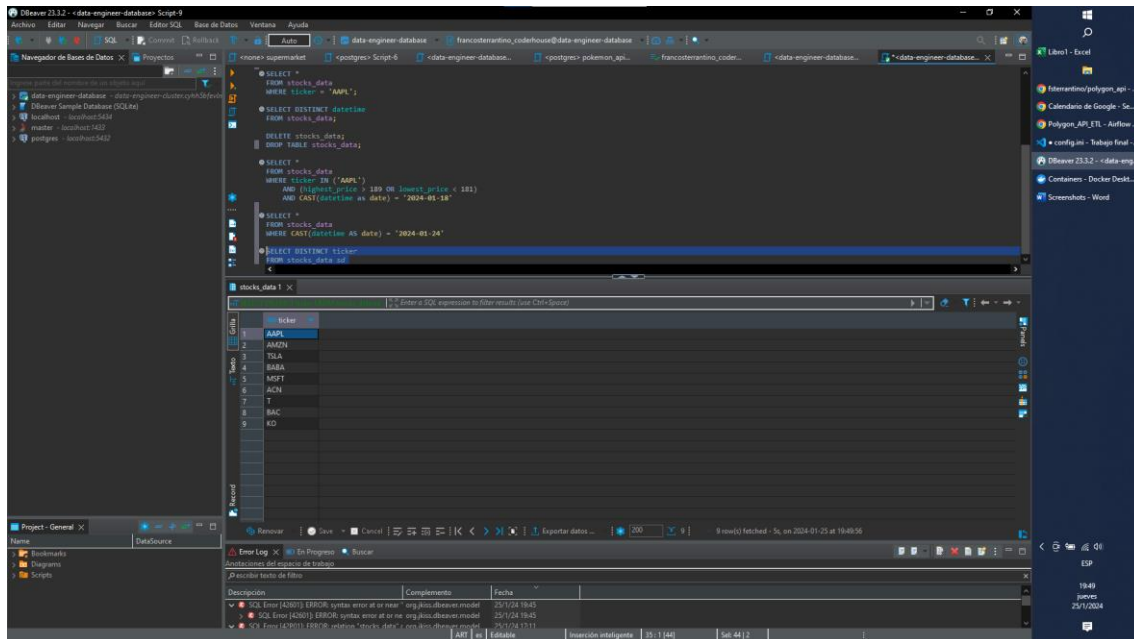
Distinct datetimes are: 22, 23, 24. As expected. We have some rows for January 24th because in the extraction of 23 it catches the Price for 24/01/2024 00:00:00.000.



January 24th screen:

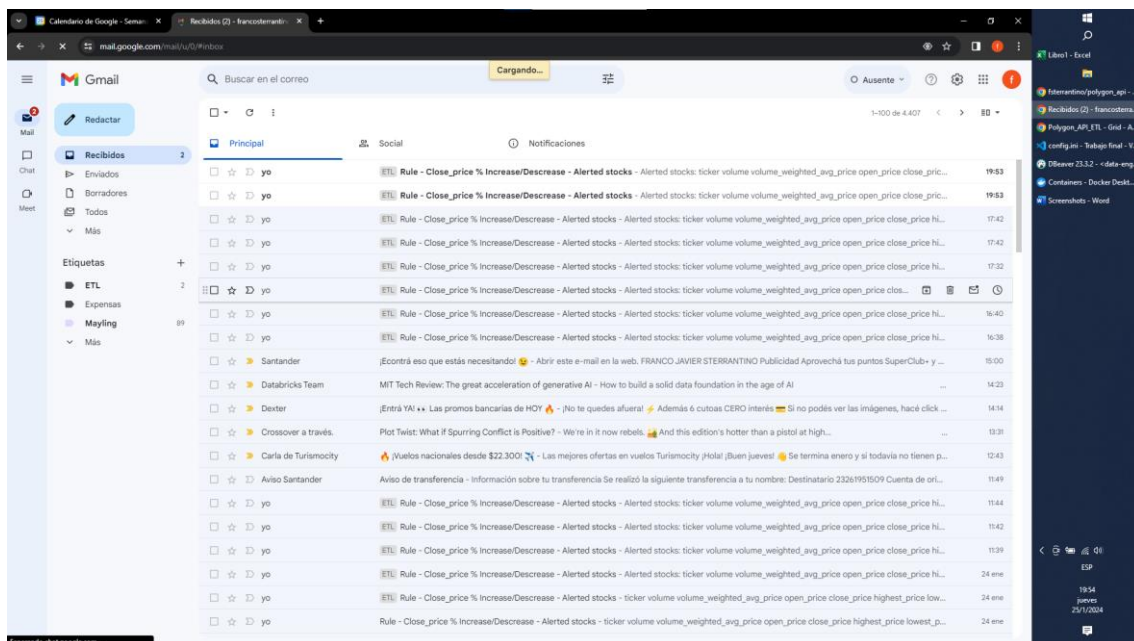


Checking distinct tickers: 9 tickers as expected.



Alerted stock: I had to “clear” the tasks because one config.ini parameter was wrong. Decrease_threshold must be lower than 0.

```
[ALERTS_PARAMS]
percentage_increase_threshold = 3
percentage_decrease_threshold = -2
```



TLSA alerted on January 23rd. The stock Price moved from 216.9475 to 210.04 (-3.183950%). More decrease than the threshold (-2%).

Rule - Close_price % Increase/Decrease - Alerted stocks

francosterrantino@gmail.com

Alerted stocks:

ticker	volume	volume_weighted_avg_price	open_price	close_price	highest_price	lowest_price	datetime	transactions_number	extraction_date	percentage_change	exceeded
37[TSLA]	1930868.0	213.4793	212.5690	216.9475	217.61	210.1190	2024-01-22 14:00:00	262877	2024-01-23	2.635321	False
38[TSLA]	27805742.0	213.2432	216.9691	210.0400	217.80	208.8501	2024-01-22 15:00:00	439300	2024-01-23	-3.183950	True

BABA alerted on January 24rd. The stock Price moved from 70.20 to 73.09 (4.116809%). More increase than the threshold (3%).

Rule - Close_price % Increase/Decrease - Alerted stocks

francosterrantino@gmail.com

Alerted stocks:

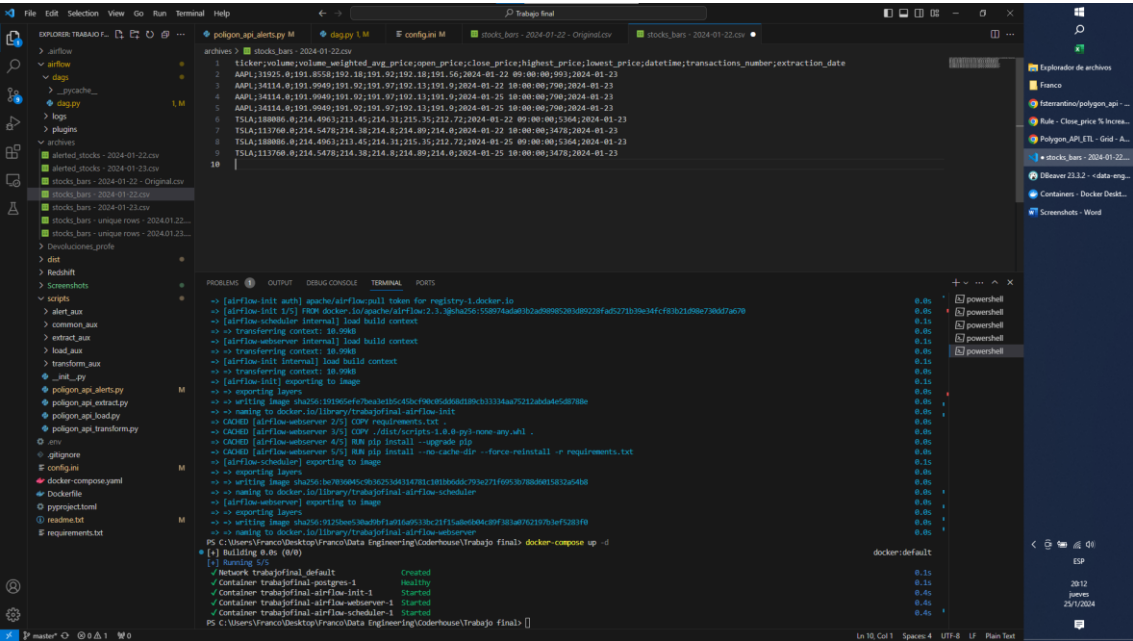
ticker	volume	volume_weighted_avg_price	open_price	close_price	highest_price	lowest_price	datetime	transactions_number	extraction_date	percentage_change	exceeded
51[BABA]	283961.0	70.0342	69.68	70.20	70.3	69.62	2024-01-23 12:00:00	3371	2024-01-24	0.762912	False
52[BABA]	3112871.0	72.3233	69.68	73.09	73.5	69.20	2024-01-23 13:00:00	27003	2024-01-24	4.116809	True

I downloaded a csv with the data loaded in the DB to check this control. The result of the script is correct:



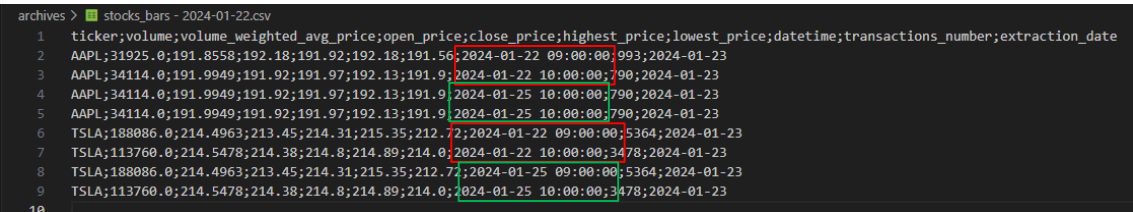
stocks_data_202401
252000.csv

Now, I would try to test the integrity control. With that objective, I modified the file that it's loaded on January 22nd.

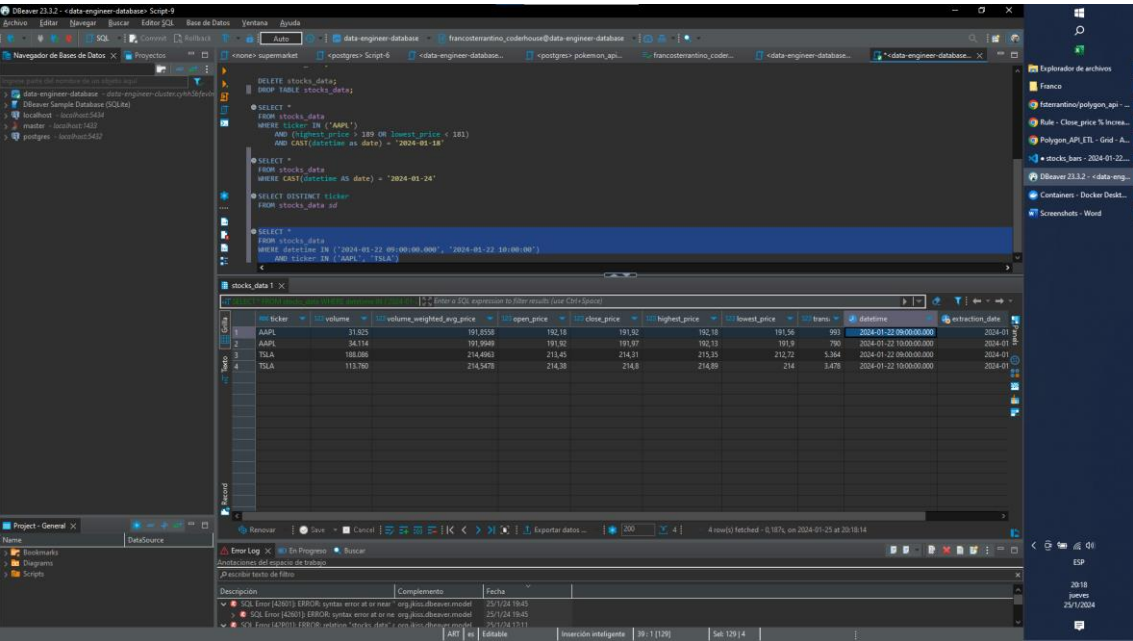


The first two rows of AAPL and TSLA, are the original ones, and they must be denied.

The other rows, are a copy with the datetime modified to 2024-01-25, in order to pass the control and be inserted.



We can check that red rows are effectively loaded right now:



I replaced the file in the Docker container:

```

C:\Users\Francisco>docker ps
Microsoft Windows [versión 10.0.19045.938]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\Francisco>docker ps
CONTAINER ID        IMAGE               COMMAND                  CREATED              STATUS              PORTS                               NAMES
b8a6f5c8406        trabaiofinal-airflow-init   "/usr/bin/dumb-init -"  8 minutes ago       Up 27 seconds      8080/tcp                           trabaiofinal-airflow-init-1
d12a2d08568        trabaiofinal-airflow-webserver "/usr/bin/dumb-init -"  8 minutes ago       Up 27 seconds (health: starting)  0.0.0.0:8080->8080/tcp             trabaiofinal-airflow-webserver-1
b15724185d2        trabaiofinal-airflow-scheduler "/usr/bin/dumb-init -"  8 minutes ago       Up 27 seconds      8080/tcp                           trabaiofinal-airflow-scheduler-1
9c1c9532428        postgres:13           "docker-entrypoint.s..."  8 minutes ago       Up 33 seconds (healthy)           0.0.0.0:5431->5432/tcp             trabaiofinal-postgres-1

C:\Users\Francisco>docker exec -it d12a2d08568 /bin/bash
airflow@b12a2d08568:/opt/archives/$
airflow@b12a2d08568:/opt/archives/$ ls
alerted_stocks - 2024-01-22.csv      alerted_stocks - 2024-01-23.csv      'stocks_bars - 2024-01-22.csv'      'stocks_bars - 2024-01-23.csv'      'stocks_bars - 2024-01-24.csv'      'stocks_bars - unique rows - 2024-01-23.csv'
alerted_stocks - 2024-01-23.csv      'stocks_bars - 2024-01-22.csv'      'stocks_bars - 2024-01-23.csv'      'stocks_bars - unique rows - 2024-01-22.csv'      'stocks_bars - unique rows - 2024-01-24.csv'
dicker;volume,volume_weighted_avg_price;open_price;close_price;datetime;transactions_number;extraction_date
AAPL:13925.0;191.8558;192.18;191.92;192.18;191.56;2024-01-22 09:00:00;991;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-22 18:00:00;798;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 18:00:00;798;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 18:00:00;798;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-22 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 10:00:00;1478;2024-01-23
airflow@b12a2d08568:/opt/archives$ cat 'stocks_bars - 2024-01-22.csv'
AAPL:13925.0;191.8558;192.18;191.92;192.18;191.56;2024-01-22 09:00:00;991;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-22 18:00:00;798;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 18:00:00;798;2024-01-23
AAPL:14114.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 18:00:00;798;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-22 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 09:00:00;1564;2024-01-23
AAPL:13760.0;214.4963;213.45;214.13;215.35;212.72;2024-01-25 10:00:00;1478;2024-01-23
airflow@b12a2d08568:/opt/archives$

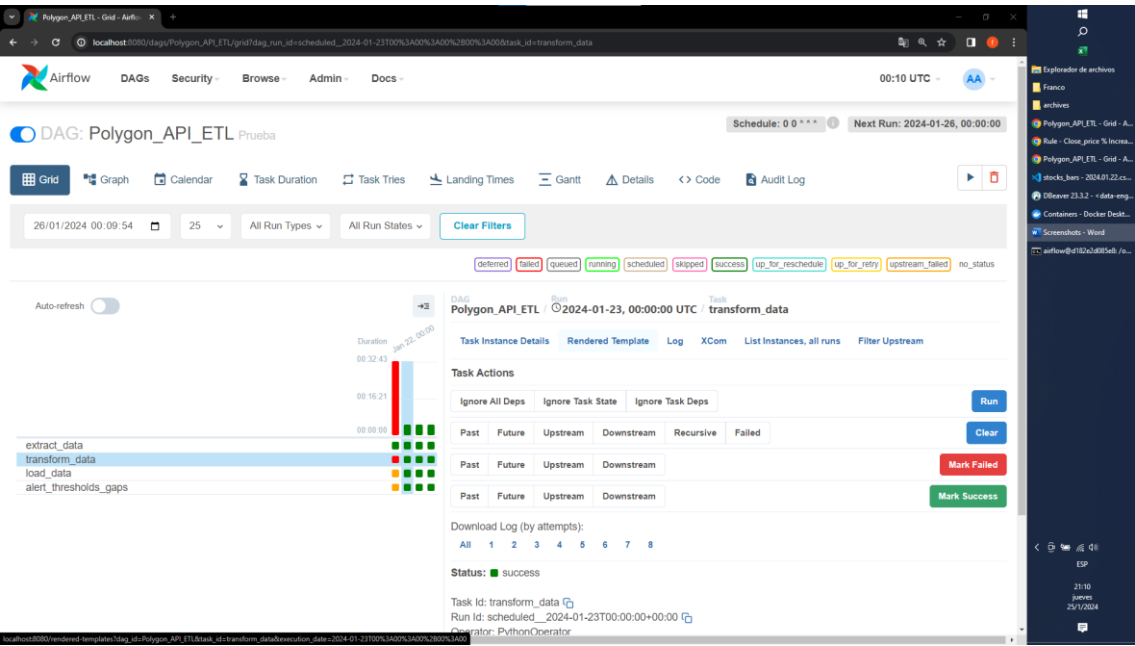
```

```
airflow@182e2d085e8:/opt/archives$
airflow@182e2d085e8:/opt/archives$ cat 'stocks_bars - 2024.01.22.csv'
ticker;volume;volume_weighted_avg_price;open_price;close_price;highest_price;lowest_price;datetime;transactions_number;extraction_date
AAPL;31925.0;191.8558;192.18;191.92;192.18;191.56;2024-01-22 09:00:00;993;2024-01-23
AAPL;34414.0;191.9949;191.92;191.97;192.13;191.9;2024-01-22 10:00:00;790;2024-01-23
AAPL;34414.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 10:00:00;790;2024-01-23
AAPL;34414.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 10:00:00;790;2024-01-23
AAPL;34414.0;191.9949;191.92;191.97;192.13;191.9;2024-01-25 10:00:00;790;2024-01-23
AAPL;188986.0;214.4963;213.45;214.31;215.35;212.72;2024-01-22 09:00:00;5364;2024-01-23
AAPL;113769.0;214.5478;214.38;214.8;214.89;214.0;2024-01-22 10:00:00;3478;2024-01-23
AAPL;188986.0;214.4963;213.45;214.31;215.35;212.72;2024-01-25 09:00:00;5364;2024-01-23
AAPL;113769.0;214.5478;214.38;214.8;214.89;214.0;2024-01-25 10:00:00;3478;2024-01-23
airflow@182e2d085e8:/opt/archives$
```

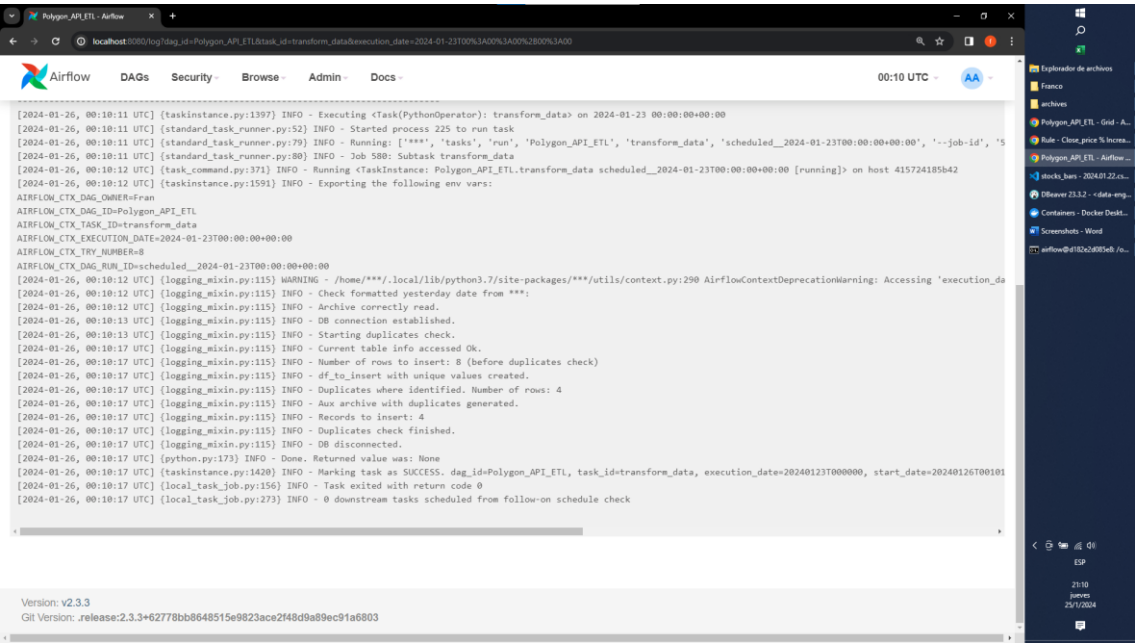
Now I'm going to clear the transform task from January 23rd (with info of 22nd) in order to re-execute it.

The screenshot displays the Airflow web interface with a modal dialog box in the foreground. The dialog, titled "Wait a minute", contains the text: "Task instances you are about to clear:" followed by a task instance ID: "TaskInstance: Polygon_API_ETL.transform_data scheduled__2024-01-23T00:00:00+00:00 [success]". At the bottom of the dialog are "Cancel" and "Confirm" buttons. In the background, the DAG "Polygon_API_ETL" is visible, showing a task list with "transform_data" highlighted. A Gantt chart on the left shows the task's duration. The interface also includes navigation tabs like "Grid", "Graph", and "Calendar", and a top navigation bar with "DAGs", "Security", "Browse", "Admin", and "Docs".

It finished with successful state:



Let's see the logs:



```

[2024-01-26, 00:10:12 UTC] {logging_mixin.py:115} WARNING - /home/***/.local/lib/python3.7/site-packages/***/utils/cont
[2024-01-26, 00:10:12 UTC] {logging_mixin.py:115} INFO - Check formatted yesterday date from ***:
[2024-01-26, 00:10:12 UTC] {logging_mixin.py:115} INFO - Archive correctly read.
[2024-01-26, 00:10:13 UTC] {logging_mixin.py:115} INFO - DB connection established.
[2024-01-26, 00:10:13 UTC] {logging_mixin.py:115} INFO - Starting duplicates check.
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Current table info accessed Ok.
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Number of rows to insert: 8 (before duplicates check)
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - df_to_insert with unique values created.
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Duplicates where identified. Number of rows: 4
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Aux archive with duplicates generated.
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Records to insert: 4
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - Duplicates check finished.
[2024-01-26, 00:10:17 UTC] {logging_mixin.py:115} INFO - DB disconnected.
[2024-01-26, 00:10:17 UTC] {python.py:173} INFO - Done. Returned value was: None
[2024-01-26, 00:10:17 UTC] {taskinstance.py:1420} INFO - Marking task as SUCCESS. dag_id=Polygon_API_ETL, task_id=trans
[2024-01-26, 00:10:17 UTC] {local_task_job.py:156} INFO - Task exited with return code 0
[2024-01-26, 00:10:17 UTC] {local_task_job.py:273} INFO - 0 downstream tasks scheduled from follow-on schedule check

```

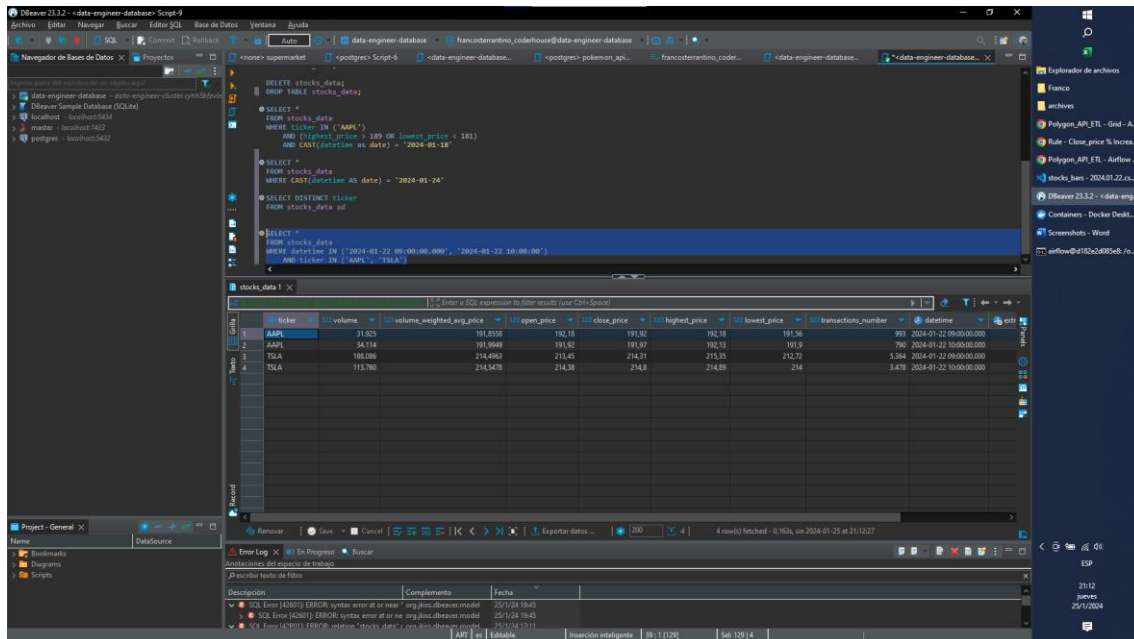
It detected successfully the 4 rows duplicated. Let's see them in the file generated:

The terminal shows the contents of a directory containing several CSV files. The files are: 'stocks_bars - 2024.01.22 - original.csv', 'stocks_bars - 2024.01.23.csv', 'stocks_bars - 2024.01.24.csv', 'stocks_bars - duplicated rows - 2024.01.22.csv', and 'stocks_bars - unique rows - 2024.01.22.csv'. The 'stocks_bars - duplicated rows - 2024.01.22.csv' file is then opened, showing a CSV with 4 duplicated rows. The CSV has columns: ticker, volume, volume_weighted_avg_price, open_price, close_price, highest_price, lowest_price, transactions_number, datetime, extraction_date, load_date, volume_weighted_avg_price, open_price, close_price, highest_price, lowest_price, transactions_number, datetime, extraction_date, load_date. The duplicated rows are for AAPL, AAPL, AAPL, and AAPL.

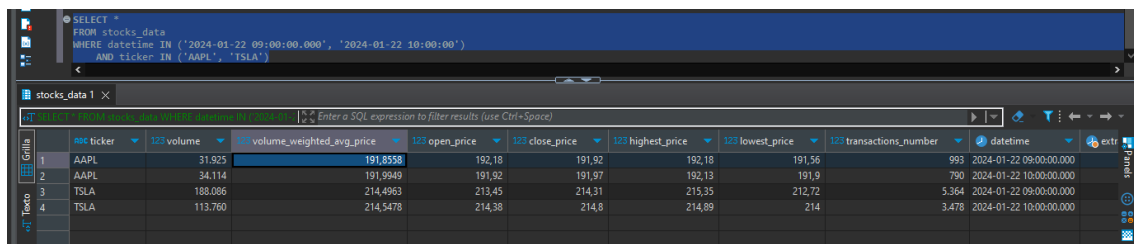
We can see the 4 duplicated rows:

The terminal shows the contents of the 'stocks_bars - duplicated rows - 2024.01.22.csv' file. The CSV has 4 duplicated rows. The rows are: AAPL;31925.0;191.8558;192.18;191.92;192.18;191.56;993.0;2024-01-22 09:00:00;2024-01-23;31925.0;191.8558;192.18;191.92;192.18;191.56;993;2024-01-23;AAPL;34114.0;191.9949;191.92;191.97;192.13;191.9;790.0;2024-01-22 10:00:00;2024-01-23;34114.0;191.9949;191.92;191.97;192.13;191.9;790;2024-01-23;AAPL;188086.0;214.4963;213.45;214.31;215.35;212.72;5364.0;2024-01-22 09:00:00;2024-01-23;188086.0;214.4963;213.45;214.31;215.35;212.72;5364;2024-01-23;AAPL;113760.0;214.5478;214.38;214.8;214.89;214.0;3478.0;2024-01-22 10:00:00;2024-01-23;113760.0;214.5478;214.38;214.8;214.89;214.0;3478;2024-01-23.

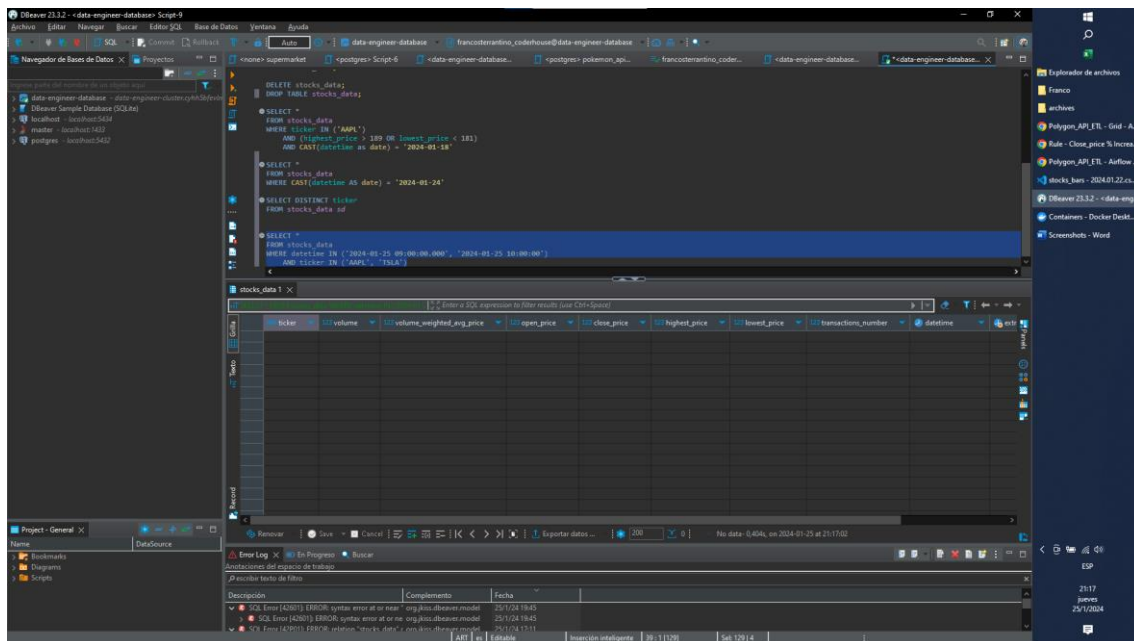
And finally we can check the DB:



The query used before continues giving us the 4 original rows:



The four additional rows are not inserted because the load task was not executed:



The screenshot displays the DBeaver IDE interface. At the top, the SQL editor contains the following query:

```
SELECT *
FROM stocks_data
WHERE datetime IN ('2024-01-25 09:00:00.000', '2024-01-25 10:00:00')
AND ticker IN ('AAPL', 'TSLA')
```

Below the editor, the table viewer for 'stocks_data 1' is shown. It has the following columns: ticker, volume, volume_weighted_avg_price, open_price, close_price, highest_price, lowest_price, transactions_number, and datetime. The table is currently empty.

The bottom status bar shows an error log with the following messages:

- SQL Error [25001]: ERROR: syntax error at or near "org.kiss.dbeaver.model"
- SQL Error [42601]: ERROR: syntax error at or near "org.kiss.dbeaver.model"
- SQL Error [42601]: ERROR: relation "stocks_data" does not exist

The right sidebar shows a file explorer with 'Screenshots - Word' and a system tray with the time '21:17' and date 'jueves 25/1/2024'.