

FINIDORI LUCA

TP_DOCKER

Etape 1 : Créer un environnement et installer le nécessaire

```
PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools> python -m venv dbt-env
PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools> dbt-env\Scripts\activate
(dbt-env) PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools> pip install dbt-core dbt-postgres
```

Etape 2 : Faire un *dbt init* et rentrer les informations

```
Enter a number: 1
host (hostname for the instance): 192.168.1.73
port [5432]: 5432
user (dev username): postgres
pass (dev password):
dbname (default database that dbt will build objects in): postgres
schema (default schema that dbt will build objects in): public
threads (1 or more) [1]: 1
```

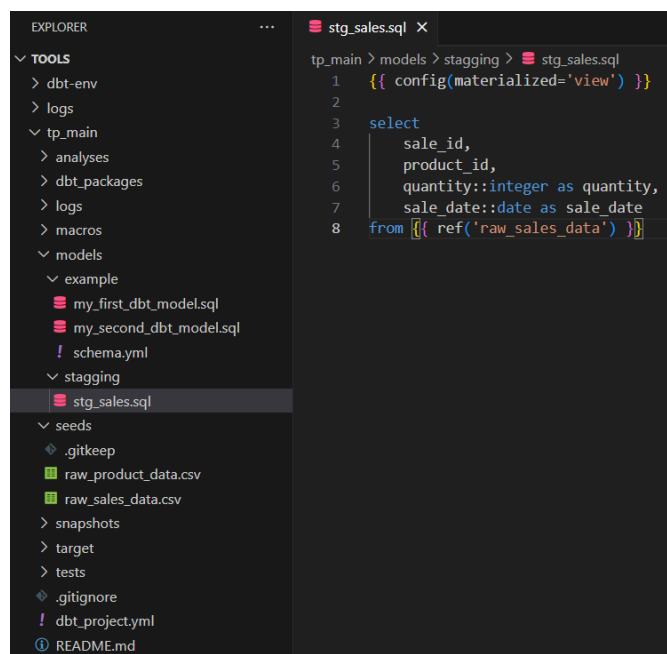
Après un *dbt debug* :

```
15:25:14 All checks passed!
```

Etape 3 : Mettre les CSV dans seed et lancer la commande dbt seed

```
(dbt-env) PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools\tp_main> dbt seed
15:30:59 Running with dbt=1.7.11
15:30:59 Registered adapter: postgres=1.7.11
15:30:59 Unable to do partial parsing because saved manifest not found. Starting full parse.
15:31:00 Found 2 models, 2 seeds, 4 tests, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:31:00 Concurrency: 1 threads (target='dev')
15:31:00 1 of 2 START seed file public.raw_product_data ..... [RUN]
15:31:01 1 of 2 OK loaded seed file public.raw_product_data ..... [INSERT 4 in 0.17s]
15:31:01 2 of 2 START seed file public.raw_sales_data ..... [RUN]
15:31:01 2 of 2 OK loaded seed file public.raw_sales_data ..... [INSERT 500 in 0.74s]
15:31:01 Finished running 2 seeds in 0 hours 0 minutes and 1.16 seconds (1.16s).
15:31:01 Completed successfully
15:31:01 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

Etape 4 : Créer un dossier Staging dans model et créer le fichier



```
EXPLORER
...
TOOLS
  > dbt-env
  > logs
  > tp_main
    > analyses
    > dbt_packages
    > logs
    > macros
    > models
      > example
        my_first_dbt_model.sql
        my_second_dbt_model.sql
        ! schema.yml
      > staging
        stg_sales.sql
    > seeds
      .gitkeep
      raw_product_data.csv
      raw_sales_data.csv
    > snapshots
    > target
    > tests
    .gitignore
    ! dbt_project.yml
    README.md

stg_sales.sql X
tp_main > models > staging > stg_sales.sql
1 {{ config(materialized='view') }}
2
3 select
4   sale_id,
5   product_id,
6   quantity::integer as quantity,
7   sale_date::date as sale_date
8 from {{ ref('raw_sales_data') }}
```

FINIDORI LUCA

TP_DOCKER

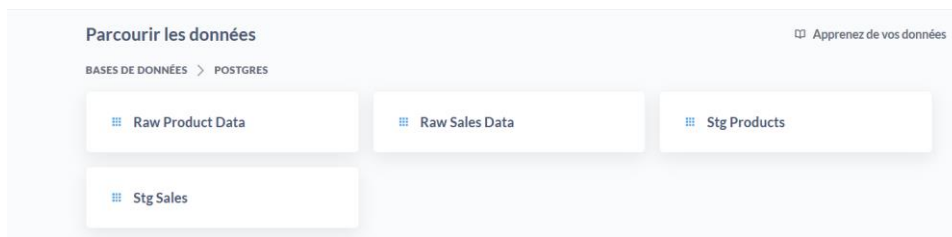
Etape 5 : Run le fichier avec la commande `dbt run --select stg_sales` :

```
(dbt-env) PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools\tp_main> dbt run --select stg_sales
15:35:01 Running with dbt=1.7.11
15:35:01 Registered adapter: postgres=1.7.11
15:35:01 Found 3 models, 2 seeds, 4 tests, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:35:01
15:35:01 Concurrency: 1 threads (target='dev')
15:35:01
15:35:01 1 of 1 START sql view model public.stg_sales ..... [RUN]
15:35:01 1 of 1 OK created sql view model public.stg_sales ..... [CREATE VIEW in 0.16s]
15:35:01
15:35:02 Finished running 1 view model in 0 hours 0 minutes and 0.37 seconds (0.37s).
15:35:02
15:35:02 Completed successfully
15:35:02
15:35:02 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

J'ai du modifier mon hôte avec 172.22.64.1, donc après la manipulation on obtient :



Etape 6 : Créer le fichier product et run le fichier avec la commande `dbt run --select stg_products` :



Etape 7 : Créer le models.yml et lancer `dbt test --select stg_sales`

```

└─ tp_main
  └─ analyses
  └─ dbt_packages
  └─ logs
  └─ macros
  └─ models
    └─ example
      └─ my_first_dbt_model.sql
      └─ my_second_dbt_model.sql
      └─ schema.yml
      └─ staging
        └─ models.yml
        └─ stg_products.sql
        └─ stg_sales.sql
    └─ seeds
    └─ snapshots
    └─ target
    └─ tests
    └─ .gitignore
    └─ dbt_project.yml
    └─ README.md

(dbt-env) PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools\tp_main> dbt test --select stg_sales
16:00:20 Running with dbt=1.7.11
16:00:21 Registered adapter: postgres=1.7.11
16:00:21 Found 4 models, 2 seeds, 6 tests, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
16:00:21
16:00:21 Concurrency: 1 threads (target='dev')
16:00:21
16:00:21 1 of 2 START test not null stg_sales_sale_id ..... [RUN]
16:00:21 1 of 2 PASS not null stg_sales_sale_id ..... [PASS in 0.09s]
16:00:21 2 of 2 START test unique stg_sales_sale_id ..... [RUN]
16:00:21 2 of 2 PASS unique stg_sales_sale_id ..... [PASS in 0.06s]
16:00:21
16:00:21 Finished running 2 tests in 0 hours 0 minutes and 0.34 seconds (0.34s).
16:00:21
16:00:21 Completed successfully
16:00:21
16:00:21 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

Etape 8: modifier la ligne 4 et 7 du models.yml avec les infos données et lancer le test

```
(dbt-env) PS C:\Users\Lucaa\Desktop\Big Data&Digital Marketing\WebService\Tools\tp_main> dbt test --select stg_products
16:06:01 Running with dbt=1.7.11
16:06:01 Registered adapter: postgres=1.7.11
16:06:01 Found 4 models, 2 seeds, 6 tests, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
16:06:01
16:06:01 Concurrency: 1 threads (target='dev')
16:06:01
16:06:01 1 of 2 START test not null stg_products_product_id ..... [RUN]
16:06:01 1 of 2 PASS not null stg_products_product_id ..... [PASS in 0.08s]
16:06:02 2 of 2 START test unique stg_products_product_id ..... [RUN]
16:06:02 2 of 2 PASS unique stg_products_product_id ..... [PASS in 0.06s]
16:06:02
16:06:02 Finished running 2 tests in 0 hours 0 minutes and 0.31 seconds (0.31s).
16:06:02
16:06:02 Completed successfully
16:06:02
16:06:02 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

Etape 9 : créer le fichier sql demandé et le lancer

```
tp_main > models > marts > daily_sales_revenue_by_category.sql
1  {{ config(materialized='table') }}
2
3  select
4      s.sale_date,
5      p.category
6      sum(s.quantity*p.price) as daily_revenue
7  from {{ ref('stg_sales') }} s
8  join {{ ref('stg_products') }} p
9      on s.product_id = p.product_id
10 group by s.sale_date , p.category
```

```
● (dbt-env) PS C:\Users\Lucaa\Desktop\Big Data\Digital Marketing\WebService\Tools\tp_main> dbt run --models daily_sales_revenue_by_category
16:22:21 Running with dbt=1.7.11
16:22:21 Registered adapter: postgres=1.7.11
16:22:21 Found 5 models, 2 seeds, 6 tests, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
16:22:21 Concurrency: 1 threads (target='dev')
16:22:21
16:22:21 1 of 1 START sql table model public.daily_sales_revenue_by_category ..... [RUN]
16:22:22 1 of 1 OK created sql table model public.daily_sales_revenue_by_category ..... [SELECT 169 in 0.19s]
16:22:22
16:22:22 Finished running 1 table model in 0 hours 0 minutes and 0.39 seconds (0.39s).
16:22:22
16:22:22 Completed successfully
16:22:22
16:22:22 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

Etape Finale : Visualiser le tout



FINIDORI LUCA
TP_DOCKER

● Gadgets 49.7%
● Widgets 50.3%

