

TP DBT

Estelle CLEMENT 2DG

1. Setting Up the DBT Project

```
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
• (base) MacBook-Air-de-Estelle:On Data science tools estelleclement$ python3 -m venv dbt-env
source dbt-env/bin/activate
• (base) MacBook-Air-de-Estelle:On Data science tools estelleclement$ source dbt-env/bin/activate
• (dbt-env) (base) MacBook-Air-de-Estelle:On Data science tools estelleclement$ source dbt-env/bin/activate
• (dbt-env) (base) MacBook-Air-de-Estelle:On Data science tools estelleclement$ pip install dbt-core dbt-postgres
Collecting dbt-core
  Using cached dbt_core-1.7.10-py3-none-any.whl.metadata (4.0 kB)
Collecting dbt-postgres
  Using cached dbt_postgres-1.7.10-py3-none-any.whl.metadata (2.8 kB)
Collecting agate~1.7.0 (from dbt-core)
  Using cached agate-1.7.1-py2.py3-none-any.whl.metadata (3.1 kB)
Collecting Jinja2<4,>=3.1.3 (from dbt-core)
  Downloading Jinja2-3.1.3-py3-none-any.whl.metadata (3.3 kB)
Collecting mashumaro~3.9 (from mashumaro[msgpack]~3.9->dbt-core)
  Using cached mashumaro-3.12-py3-none-any.whl.metadata (111 kB)
Collecting logbook<1.6,>=1.5 (from dbt-core)
  Using cached Logbook-1.5.3.tar.gz (85 kB)
```

```
Downloading idna-3.6-py3-none-any.whl (61 kB)
 61.6/61.6 kB 2.1 MB/s eta 0:00:00
Using cached isodate-0.6.1-py2.py3-none-any.whl (41 kB)
Downloading Jinja2-3.1.3-py3-none-any.whl (133 kB)
 133.2/133.2 kB 2.6 MB/s eta 0:00:00
Downloading jsonschema-4.21.1-py3-none-any.whl (85 kB)
 85.5/85.5 kB 2.3 MB/s eta 0:00:00
Using cached mashumaro-3.12-py3-none-any.whl (89 kB)
Downloading networkx-3.2.1-py3-none-any.whl (1.6 MB)
 1.6/1.6 MB 4.9 MB/s eta 0:00:00
Downloading packaging-24.0-py3-none-any.whl (53 kB)
 53.5/53.5 kB 2.3 MB/s eta 0:00:00
Downloading pathspec-0.11.2-py3-none-any.whl (29 kB)
Using cached protobuf-4.25.3-cp37-abi3-macosx_10_9_universal2.whl (394 kB)
Downloading pycpg2_binary-2.9.9-cp312-cp312-macosx_10_9_x86_64.whl (2.8 MB)
 2.8/2.8 MB 7.8 MB/s eta 0:00:00
Downloading pytz-2024.1-py2.py3-none-any.whl (505 kB)
 505.5/505.5 kB 7.6 MB/s eta 0:00:00
Downloading PyYAML-6.0.1-cp312-cp312-macosx_10_9_x86_64.whl (178 kB)
 178.7/178.7 kB 6.1 MB/s eta 0:00:00
Downloading requests-2.31.0-py3-none-any.whl (62 kB)
```

Configure dbt profile

```
• (dbt-env) (base) MacBook-Air-de-Estelle:On Data science tools estelleclement$ dbt init
13:43:42 Running with dbt=1.7.10
13:43:42 [ConfigFolderDirectory]: Unable to parse dict {'dir': PosixPath('/Users/estelleclement/.dbt')}
13:43:42 Creating dbt configuration folder at
Enter a name for your project (letters, digits, underscore): TP dbt
TP dbt is not a valid project name.
Enter a name for your project (letters, digits, underscore): tp_dbt
13:44:25
Your new dbt project "tp_dbt" was created!

For more information on how to configure the profiles.yml file,
please consult the dbt documentation here:

https://docs.getdbt.com/docs/configure-your-profile

One more thing:

Need help? Don't hesitate to reach out to us via GitHub issues or on Slack:

https://community.getdbt.com/

Happy modeling!

13:44:25 Setting up your profile.
Which database would you like to use?
[1] postgres
```

```

Enter a number: 1
host (hostname for the instance): 10.4.251.244
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
14:03:24 Profile tp_dbt written to /Users/estelleclement/.dbt/profiles.yml using target's profile_template.yml and your supplied values. Run 'dbt debug' to
validate the connection.

```

```

• (dbt-env) (base) MacBook-Air-de-Estelle:0n Data science tools estelleclement$ ls
TP.py dbt-env logs tp_dbt
• (dbt-env) (base) MacBook-Air-de-Estelle:0n Data science tools estelleclement$ cd tp_dbt
• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt debug
14:03:57 Running with dbt=1.7.10
14:03:57 dbt version: 1.7.10
14:03:57 python version: 3.12.2
14:03:57 python path: /Users/estelleclement/Desktop/ESME/Ingé 2/0n Data science tools/dbt-env/bin/python3.12
14:03:57 os info: macOS-12.7.3-x86_64-i386-64bit
14:03:57 Using profiles dir at /Users/estelleclement/.dbt
14:03:57 Using profiles.yml file at /Users/estelleclement/.dbt/profiles.yml
14:03:57 Using dbt_project.yml file at /Users/estelleclement/Desktop/ESME/Ingé 2/0n Data science tools/tp_dbt/dbt_project.yml
14:03:57 adapter type: postgres
14:03:57 adapter version: 1.7.10
14:03:57 Configuration:
14:03:57   profiles.yml file [OK found and valid]
14:03:57   dbt_project.yml file [OK found and valid]
14:03:57 Required dependencies:
14:03:57   - git [OK found]

14:03:57 Connection:
14:03:57   host: 10.4.251.244
14:03:57   port: 5432
14:03:57   user: postgres
14:03:57   database: postgres
14:03:57   schema: public
14:03:57   connect_timeout: 10
14:03:57   role: None
14:03:57   search_path: None
14:03:57   keepalives_idle: 0
14:03:57   sslmode: None
14:03:57   sslcert: None
14:03:57   sslkey: None
14:03:57   sslrootcert: None
14:03:57   application_name: dbt
14:03:57   retries: 1
14:03:57 Registered adapter: postgres=1.7.10
14:03:57 Connection test: [OK connection ok]

14:03:57 All checks passed!

```

2. Preparing the Raw data

```

• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt seed
14:18:17 Running with dbt=1.7.10
14:18:18 Registered adapter: postgres=1.7.10
14:18:18 Found 3 models, 4 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
14:18:18 Concurrency: 1 threads (target='dev')
14:18:19
14:18:19 1 of 2 START seed file public.raw_product_data ..... [RUN]
14:18:19 1 of 2 OK loaded seed file public.raw_product_data ..... [INSERT 4 in 0.23s]
14:18:19 2 of 2 START seed file public.raw_sales_data ..... [RUN]
14:18:20 2 of 2 OK loaded seed file public.raw_sales_data ..... [INSERT 500 in 1.20s]
14:18:20
14:18:20 Finished running 2 seeds in 0 hours 0 minutes and 1.89 seconds (1.89s).
14:18:20 Completed successfully
14:18:20
14:18:20 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2

```

3. Creating Staging Models

3.1 Staging Sales Data (stg_sales.sql)

```
• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt run --select stg_sales
14:22:32 Running with dbt=1.7.10
14:22:33 Registered adapter: postgres=1.7.10
14:22:33 Found 3 models, 4 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
14:22:33 Concurrency: 1 threads (target='dev')
14:22:33 1 of 1 START sql view model public.stg_sales ..... [RUN]
14:22:33 1 of 1 OK created sql view model public.stg_sales ..... [CREATE VIEW in 0.26s]
14:22:33 Finished running 1 view model in 0 hours 0 minutes and 0.57 seconds (0.57s).
14:22:34 Completed successfully
14:22:34 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

3.2 Staging Product Data (stg_products.sql)

```
• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt run --select stg_products
15:02:48 Running with dbt=1.7.10
15:02:48 Registered adapter: postgres=1.7.10
15:02:49 Found 4 models, 4 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:02:49 Concurrency: 1 threads (target='dev')
15:02:49 1 of 1 START sql view model public.stg_products ..... [RUN]
15:02:50 1 of 1 OK created sql view model public.stg_products ..... [CREATE VIEW in 0.43s]
15:02:50 Finished running 1 view model in 0 hours 0 minutes and 1.00 seconds (1.00s).
15:02:50 Completed successfully
15:02:50 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

3.3 Testing Staging Models

```
• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt test --select stg_sales
15:06:37 Running with dbt=1.7.10
15:06:38 Registered adapter: postgres=1.7.10
15:06:38 Found 4 models, 6 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:06:38 Concurrency: 1 threads (target='dev')
15:06:39 1 of 2 START test not_null_stg_sales_sale_id ..... [RUN]
15:06:39 1 of 2 PASS not_null_stg_sales_sale_id ..... [PASS in 0.17s]
15:06:39 2 of 2 START test unique_stg_sales_sale_id ..... [RUN]
15:06:39 2 of 2 PASS unique_stg_sales_sale_id ..... [PASS in 0.14s]
15:06:39 Finished running 2 tests in 0 hours 0 minutes and 0.62 seconds (0.62s).
15:06:39 Completed successfully
15:06:39 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

```
• (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt test --select stg_products
15:10:38 Running with dbt=1.7.10
15:10:39 Registered adapter: postgres=1.7.10
15:10:40 Found 4 models, 8 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:10:40 Concurrency: 1 threads (target='dev')
15:10:40 1 of 2 START test not_null_stg_products_product_id ..... [RUN]
15:10:40 1 of 2 PASS not_null_stg_products_product_id ..... [PASS in 0.16s]
15:10:40 2 of 2 START test unique_stg_products_product_id ..... [RUN]
15:10:41 2 of 2 PASS unique_stg_products_product_id ..... [PASS in 0.08s]
15:10:41 Finished running 2 tests in 0 hours 0 minutes and 0.57 seconds (0.57s).
15:10:41 Completed successfully
15:10:41 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

4. Marts Models

4.2 Example

```
● (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt run --select daily_sales_revenue_by_category
15:34:56 Running with dbt=1.7.10
15:34:57 Registered adapter: postgres=1.7.10
15:34:57 Found 6 models, 8 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:34:57
15:34:58 Concurrency: 1 threads (target='dev')
15:34:58
15:34:58 1 of 1 START sql table model public.daily_sales_revenue_by_category ..... [RUN]
15:34:58 1 of 1 OK created sql table model public.daily_sales_revenue_by_category ..... [SELECT 169 in 0.35s]
15:34:58
15:34:58 Finished running 1 table model in 0 hours 0 minutes and 0.73 seconds (0.73s).
15:34:58
15:34:58 Completed successfully
15:34:58
15:34:58 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

4.2 Exercice

```
● (dbt-env) (base) MacBook-Air-de-Estelle:tp_dbt estelleclement$ dbt run --select daily_sales_volume
15:43:54 Running with dbt=1.7.10
15:43:55 Registered adapter: postgres=1.7.10
15:43:55 Found 6 models, 8 tests, 2 seeds, 0 sources, 0 exposures, 0 metrics, 401 macros, 0 groups, 0 semantic models
15:43:55
15:43:56 Concurrency: 1 threads (target='dev')
15:43:56
15:43:56 1 of 1 START sql table model public.daily_sales_volume ..... [RUN]
15:43:56 1 of 1 OK created sql table model public.daily_sales_volume ..... [SELECT 90 in 0.33s]
15:43:56
15:43:56 Finished running 1 table model in 0 hours 0 minutes and 0.79 seconds (0.79s).
15:43:56
15:43:56 Completed successfully
15:43:56
15:43:56 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```