

# dbt Intro

Alexander Matrunich  
Data analyst  
HomeBuddy.com

PGUG.EE  
December 2025  
meetup

# About speaker

Degree in Sociology (Pskov Volny Institute) ->  
Surveys and opinion polls, analytics with SPSS ->  
Interest to open source ->  
Linux, R ->  
From surveys to pure data analytics ->  
Contracts with the UN FAO in Rome (R, Oracle) ->  
Data Analyst, Taxify/Bolt (Redshift, Looker, R) ->  
Data Analyst, HomeBuddy (Snowflake, dbt, Superset, Thoughtspot)

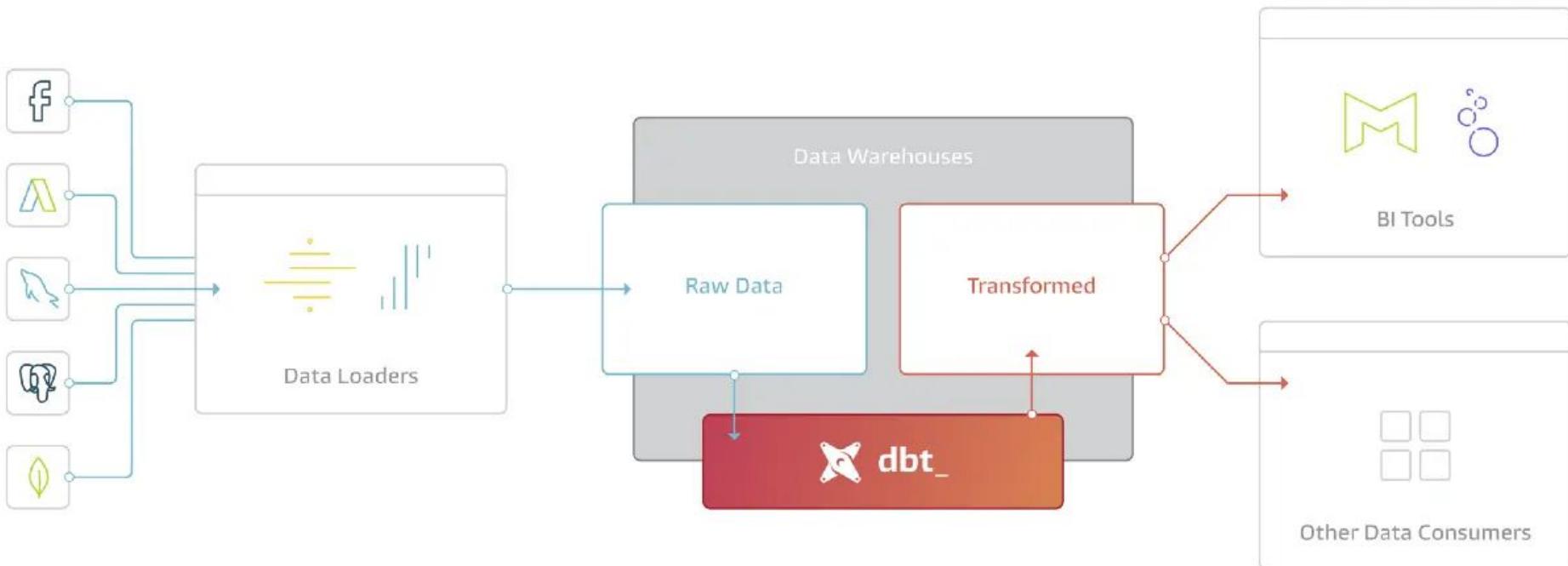
# About speaker

Degree in Sociology (Pskov Volny Institute) ->  
Surveys and opinion polls, analytics with SPSS ->  
Interest to open source ->  
Linux, R ->  
From surveys to pure data analytics ->  
Contracts with the UN FAO in Rome (R, Oracle) ->  
Data Analyst, Taxify/Bolt (Redshift, Looker, R) ->  
Data Analyst, HomeBuddy (Snowflake, dbt, Superset, Thoughtspot)

Other interests:

Paragliding (no engine/propeller!)  
Family history research

# dbt and the modern BI stack



# A dbt project

dbt\_project.yml

Contents History Compare Blame

You updated DWH\_data\_pipeline 34m ago

Create a pull request

```
1 # Name your project! Project names should contain only lowercase characters
2 # and underscores. A good package name should reflect your organization's
3 # name or the intended use of these models
4 name: 'DWH'
5 version: '1.0.0'
6 config-version: 2
7
8 # This setting configures which "profile" dbt uses for this project.
9 profile: 'azure-databricks-workspace'
10
11 # These configurations specify where dbt should look for different types of files.
12 # The 'source-paths' config, for example, states that models in this project can be
13 # found in the "models/" directory. You probably won't need to change these!
14 source-paths: ["models"]
15 analysis-paths: ["analysis"]
16 test-paths: ["tests"]
17 data-paths: ["data"]
18 macro-paths: ["macros"]
19 snapshot-paths: ["snapshots"]
20
21
22 target-path: "target" # directory which will store compiled SQL files
23 clean-targets: # directories to be removed by 'dbt clean'
24   - "target"
25   - "dbt_modules"
26
27
28 # Configuring models
29 # Full documentation: https://docs.getdbt.com/docs/configuring-models
30
31 # In this example config, we tell dbt to build all models in the example/ directory
32 # as tables. These settings can be overridden in the individual model files
33 # using the `{{ config(...) }}` macro.
34 models:
35   DWH:
36     staging:
37       +materialized: view
38       marts:
```

# A dbt model

```
1  {{{
2    config(
3      unique_key="BK_ORDER_NUMBER|| '-' || BK_ORDER_LINE_NUMBER",
4      file_format='delta'
5    )
6  }}}
7
8  with stg_sales as (
9    select * from {{ref('stg_sales')}}
10 ),
11
12 sales as (
13   SELECT
14     BK_ORDER_NUMBER,
15     BK_ORDER_LINE_NUMBER,
16     to_timestamp(OrderDateKey, 'yyyyMMdd') as ORDER_DATE,
17     M_ORDER_QUANTITY,
18     SHA1(ResellerKey) as A_HASHED_KEY_BK
19   FROM stg_sales
20 )
21
22 select * from sales
23
24 {% if is_incremental() %}
25
26   -- this filter will only be applied on an incremental run
27   where ORDER_DATE >= (select date_add(max(ORDER_DATE), -3) from {{ this }})
28
29 {% endif %}
```

```

1  {{{
2    config(
3      unique_key="BK_ORDER_NUMBER|| '-' || BK_ORDER_LINE_NUMBER",
4      file_format='delta'
5    )
6  }}}
7
8  with stg_sales as (
9    select * from {{ref('stg_sales')}}
10 ),
11
12 sales as (
13   SELECT
14     BK_ORDER_NUMBER,
15     BK_ORDER_LINE_NUMBER,
16     to_timestamp(OrderDateKey, 'yyyyMMdd') as ORDER_DATE,
17     M_ORDER_QUANTITY,
18     SHA1(ResellerKey) as A_HASHED_KEY_BK
19   FROM stg_sales
20 )
21
22 select * from sales
23
24 {% if is_incremental() %}
25
26   -- this filter will only be applied on an incremental run
27   where ORDER_DATE >= (select date_add(max(ORDER_DATE), -3) from {{ this }})
28
29 {% endif %}

```

# A dbt model

```

1  version: 2
2
3  models:
4    - name: stg_customers
5      columns:
6        - name: BK_CUSTOMER_CODE
7          tests:
8            - unique
9            - not_null
10   - name: dim_customers
11     columns:
12       - name: BK_CUSTOMER_CODE
13         tests:
14           - unique
15           - not_null

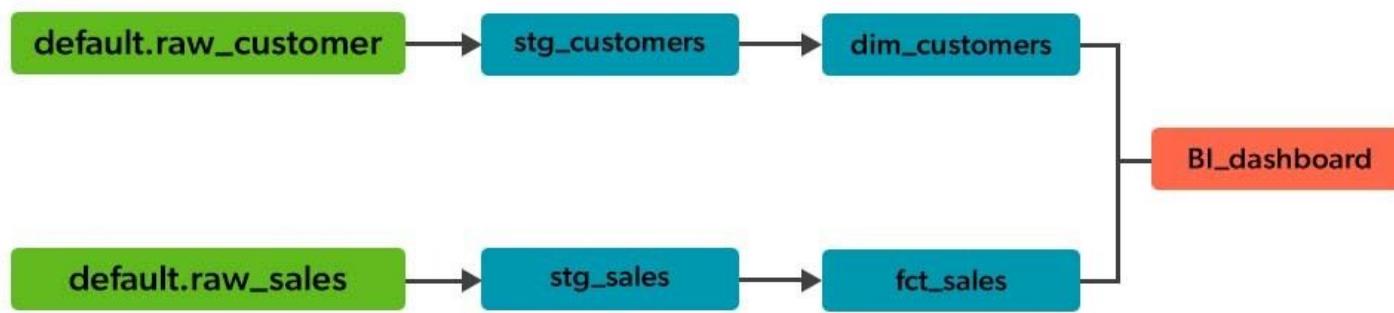
```

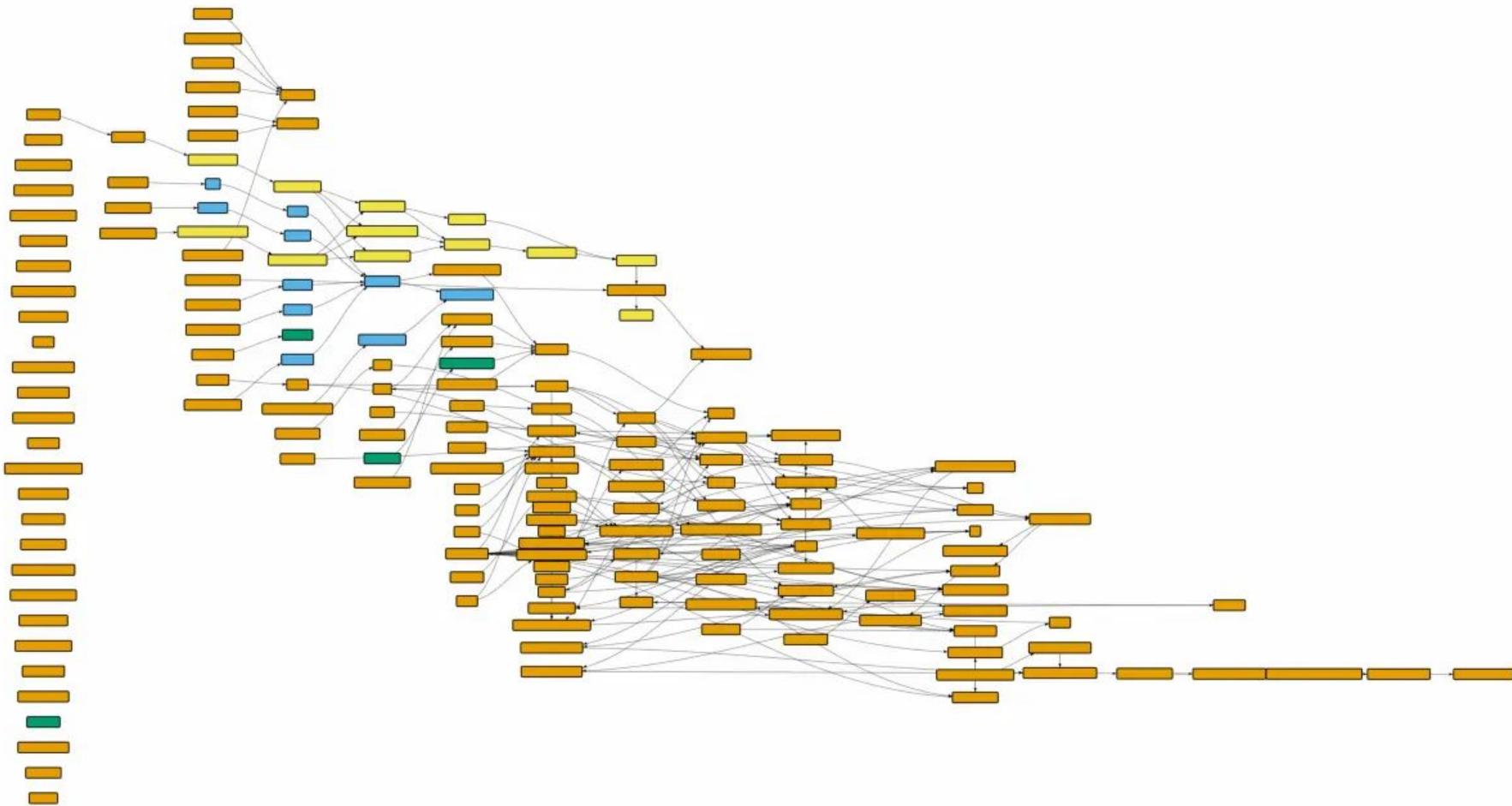
```
$ cat -n models/orders_rollup.sql
 1 with
 2   orders as ( select * from {{ref('orders')}} )
 3 , line_items as ( select * from {{ref('line_items')}} )
 4
 5 select
 6   id
 7 , sum(line_items.purchase_price)
 8
 9 from orders
10 left join line_items
11   on orders.id = line_items.order_id
12
13 group by 1
$ dbt run
Found 3 models, 0 tests, 0 archives, 0 analyses, 10 macros, 0 operations

09:44:39 | Concurrency: 4 threads (target='postgres')
09:44:39 |
09:44:39 | 1 of 3 START view model ecom.orders..... [RUN]
09:44:39 | 2 of 3 START view model ecom.line_items..... [RUN]
09:44:39 | 1 of 3 OK created view model ecom.orders..... [CREATE VIEW in 0.07s]
09:44:39 | 2 of 3 OK created view model ecom.line_items..... [CREATE VIEW in 0.07s]
09:44:39 | 3 of 3 START view model ecom.orders_rollup..... [RUN]
09:44:39 | 3 of 3 OK created view model ecom.orders_rollup..... [CREATE VIEW in 0.05s]
09:44:39 |
09:44:39 | Finished running 3 view models in 0.43s.
```

Completed successfully

# DAG (Directed Acyclic Graph)





# Jinja, a templating language

```
select * from {{ref('really_big_table')}}
```

```
{%raw%}
```

```
{% if incremental and target.schema == 'prod' %}
```

```
    where timestamp >= (select max(timestamp) from {{this}})
```

```
{% else %}
```

```
    where timestamp >= dateadd(day, -3, current_date)
```

```
{% endif %}
```

README Contributing Apache-2.0 license

- [Installation Instructions](#)

- [Generic Tests](#)

- [equal\\_rowcount \(source\)](#)
- [fewer\\_rows\\_than \(source\)](#)
- [equality \(source\)](#)
- [expression\\_is\\_true \(source\)](#)
- [recency \(source\)](#)
- [at\\_least\\_one \(source\)](#)
- [not\\_constant \(source\)](#)
- [not\\_empty\\_string \(source\)](#)
- [cardinality\\_equality \(source\)](#)
- [not\\_null\\_proportion \(source\)](#)
- [not\\_accepted\\_values \(source\)](#)
- [relationships\\_where \(source\)](#)
- [mutually\\_exclusive\\_ranges \(source\)](#)
- [sequential\\_values \(source\)](#)
- [unique\\_combination\\_of\\_columns \(source\)](#)
- [accepted\\_range \(source\)](#)
- [Grouping in tests](#)

- [Macros](#)

- [Introspective macros](#)

- [get\\_column\\_values \(source\)](#)
- [get\\_filtered\\_columns\\_in\\_relation \(source\)](#)
- [get\\_relations\\_by\\_pattern \(source\)](#)
- [get\\_relations\\_by\\_prefix \(source\)](#)
- [get\\_query\\_results\\_as\\_dict \(source\)](#)
- [get\\_single\\_value \(source\)](#)

- [SQL generators](#)

- [date\\_spine \(source\)](#)
- [deduplicate \(source\)](#)
- [haversine\\_distance \(source\)](#)
- [group\\_by \(source\)](#)
- [star \(source\)](#)
- [union\\_relations \(source\)](#)
- [generate\\_series \(source\)](#)
- [generate\\_surrogate\\_key \(source\)](#)
- [safe\\_add \(source\)](#)
- [safe\\_divide \(source\)](#)
- [safe\\_subtract \(source\)](#)
- [pivot \(source\)](#)
- [unpivot \(source\)](#)
- [width\\_bucket \(source\)](#)

- [Web macros](#)

# dbt alternatives?

dbt

## Software Development practices for Analytics:

- Source control
- Tests
- Documentation
- Development environment