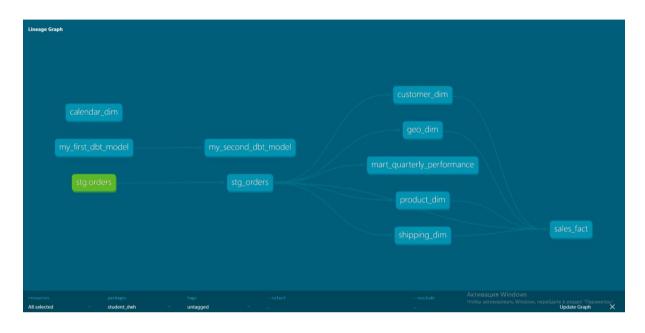
Проект: student\_dwh

Выполнил(а): Бурлов Василий Тимофеевич

Группа: БД-251м

#### Граф зависимостей:



## Код модели stg\_orders.sql:

- -- models/staging/stg\_orders.sql
- -- Эта модель читает данные из исходной таблицы stg.orders,
- -- приводит их к нужным типам и исправляет ошибку с почтовым кодом.
- -- Все последующие модели будут ссылаться на эту, а не на исходную таблицу.

#### **SELECT**

```
-- Приводим все к нижнему регистру для консистентности в dbt "order_id",
("order_date")::date as order_date,
("ship_date")::date as ship_date,
"ship_mode",
"customer_id",
"customer_name",
"segment",
"country",
"city",
"state",
-- Исправляем проблему с Burlington прямо здесь, один раз и навсегда CASE
WHEN "city" = 'Burlington' AND "postal_code" IS NULL THEN '05401'
```

```
ELSE "postal_code"
END as postal_code,
"region",
"product_id",
"category",
"subcategory" as sub_category, -- переименовываем для соответствия
"product_name",
"sales",
"quantity",
"discount",
"profit"
FROM {{ source('stg', 'orders') }}
```

## Код модели sales\_fact.sql:

```
-- Создает таблицу фактов, объединяя все измерения
SELECT
  -- Суррогатные ключи из измерений
  cd.cust_id,
  pd.prod_id,
  sd.ship_id,
  gd.geo id,
  -- Ключи для календаря
  to_char(o.order_date, 'yyyymmdd')::int AS order_date_id,
  to_char(o.ship_date, 'yyyymmdd')::int AS ship_date_id,
  -- Бизнес-ключ и метрики
  o.order id,
  o.sales,
  o.profit,
  o.quantity,
  o.discount
FROM {{ ref('stg_orders') }} AS o
LEFT JOIN {{ ref('customer_dim') }} AS cd ON o.customer_id = cd.customer_id
LEFT JOIN {{ ref('product_dim') }} AS pd ON o.product_id = pd.product_id
LEFT JOIN {{ ref('shipping_dim') }} AS sd ON o.ship_mode = sd.ship_mode
LEFT JOIN {{ ref('geo_dim') }} AS gd ON o.postal_code = gd.postal_code AND o.city =
gd.city AND o.state = gd.state
```

Код моей индивидуальной mart-модели (mart\_quarterly\_performance - вариант 5):

```
-- models/marts/mart_quarterly_performance.sql
with orders as (
  select *
  from {{ ref('stg_orders') }}
)
select
  date_trunc('quarter', order_date) as quarter,
  sum(sales) as total_revenue,
  sum(profit) as total_profit
from orders
group by 1
order by 1
```

## Файл schema.yml с тестами для всех моделей:

```
# Путь к файлу: models/marts/schema.yml
version: 2
models:
 - name: shipping_dim
  columns:
   - name: ship_id
    tests:
      - unique
      - not_null
 - name: customer_dim
  columns:
   - name: cust_id
    tests:
      - unique
      - not_null
 - name: geo_dim
  columns:
   - name: geo_id
    tests:
      - unique
      not_null
 - name: product_dim
```

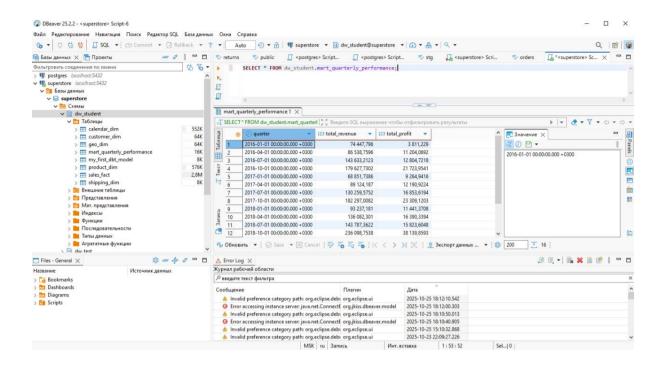
```
columns:
  - name: prod_id
   tests:
    - unique
    - not_null
- name: sales_fact
 columns:
  - name: cust_id
   tests:
     - relationships:
       arguments:
        to: ref('customer_dim')
        field: cust_id
- name: mart_quarterly_performance
 description: "Суммарная выручка и прибыль по кварталам"
 columns:
  - name: quarter
   tests:
    - not_null
    - unique
  - name: total_revenue
   tests:
     not_null
  - name: total_profit
   tests:
     - not_null
```

Скриншот успешного выполнения dbt run и dbt test для проекта student\_dwh:

```
lbt-env) C:\Users\User\Downloads\Marистратура\учеба\Платформы Data Engineering\pde_magistr\student_dwh>dbt run
14:30:37 Running with dbt=1.10.13
14:30:38 Registered adapter: postgres=1.9.1
14:30:40 Found 10 models, 17 data tests, 1 source, 448 macros
14.39.49
14:30:40 Concurrency: 1 threads (target='dev')
14:30:40
T 7670 in 1.40s]
                                                                             T 2 in 0.66sl
                                                                             E VIEW in 0.37s1
                                                                        [CRUN]
[CREATE VIEW in 0.295]
                                                                             T 16 in 0.41sl
T 4344 in 0.35s]
                                                                        [RUN]
                                                                              4 in 0.29s]
       14:30:46
14:30:47
14:30:47
14:30:47 Finished running 8 table models, 2 view models in 0 hours 0 minutes and 6.71 seconds (6.71s). 14:30:47
4:30:47 Completed successfully
4:30:47 Done. PASS=10 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=10
```

```
C:\Users\User\Downloads\Marистратура\учеба\Платформы Data Engineering\pde_magistr\student_dwh>dbt test
              Running with dbt=1.10.13
14:33:52
14:33:53
             Registered adapter: postgres=1.9.1
Found 10 models, 17 data tests, 1 source, 448 macros
14:33:53 Concurrency: 1 threads (target='dev')
14:33:54
1 of 17 START test not_null_customer_dim_cust_id
14:33:54
1 of 17 PASS not_null_customer_dim_cust_id
14:33:54
2 of 17 START test not_null_geo_dim_geo_id
14:33:54
2 of 17 PASS not_null_geo_dim_geo_id
14:33:54
3 of 17 START test not_null_mart_quarterly_performance_quarter
                                                                                                                                           [RUN]
                                                                                                                                                     in 0.23s]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.1251
in 0.23s]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.12s]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.13s]
                                                                                                                                            [RUN]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.11sl
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.11sl
                                                                                                                                            [RUN]
                                                                                                                                             [RUN]
                                                                                                                                                     in 0.15s]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.14sl
                                                                                                                                             RUN1
                                                                                                                                                     in 0.12sl
                                                                                                                                            [RUN]
             13 of 17 START test unique_mart_quarterly_performance_quarter
13 of 17 PASS unique_mart_quarterly_performance_quarter
14 of 17 START test unique_my_first_dbt_model_id
15 of 17 PASS unique_my_first_dbt_model_id
15 of 17 START test unique_my_second_dbt_model_id
15 of 17 PASS unique_my_second_dbt_model_id
16 of 17 START test unique_product_dim_prod_id
16 of 17 PASS unique_product_dim_prod_id
17 of 17 START test unique_shipping_dim_ship_id
17 of 17 PASS unique_shipping_dim_ship_id
14:33:56
14:33:56
                                                                                                                                                     in 0.11s]
                                                                                                                                            [RUN]
                                                                                                                                                     in 0.11s]
                                                                                                                                            [RUN]
 14:33:56
                                                                                                                                                     in 0.16s]
 14:33:56
                                                                                                                                            [RUN]
 14:33:56
                                                                                                                                                     in 0.1751
                                                                                                                                            [PASS
[RUN]
 14:33:56
                                                                                                                                                    in 0.12sl
 4:33:56
14:33:56 Finished running 17 data tests in 0 hours 0 minutes and 3.08 seconds (3.08s).
14:33:57
 14:33:57
 4:33:57
                  Got 1 result, configured to fail if != 0
 14:33:57
                  compiled code at target\compiled\student_dwh\models\example\schema.yml\not_null_my_first_dbt_model_id.sql
              Done. PASS=16 WARN=0 ERROR=1 SKIP=0 NO-OP=0 TOTAL=17
```

# Скриншот с данными из моей индивидуальной mart-модели (DBeaver):



#### Выводы:

Использование dbt для реализации DWH имеет несколько ключевых преимуществ по сравнению с ручным написанием DDL/DML скриптов:

**Структурированность и повторяемость**: dbt позволяет организовать проект в виде моделей, источников и макросов, что упрощает поддержку и масштабирование.

**Автоматизация тестирования данных**: встроенные тесты (unique, not\_null, relationships) помогают проверять качество данных на каждом этапе, уменьшая риск опибок.

**Документирование и lineage**: dbt автоматически генерирует документацию и граф зависимостей моделей, что облегчает понимание потока данных и коммуникацию с командой.

**Модульность**: модели можно легко комбинировать для разных аналитических задач, без дублирования SQL-кода.

**Ускорение разработки**: разработка через dbt снижает количество ручной работы, ускоряет развертывание новых моделей и улучшает контроль версий с помощью Git.

В целом, dbt делает процесс построения DWH более прозрачным, управляемым и надежным по сравнению с ручным созданием таблиц и написанием SQL-скриптов.