# Департамент образования и науки города Москвы

Государственное автономное образовательное учреждение высшего образования города Москвы «Московский городской педагогический университет»

Институт цифрового образования

Лабораторная работа 2.1 по дисциплине «Платформы Data Engineering»

Выполнил: студент БД-251м

Быков Владимир

### 1. Краткое описание архитектуры проекта

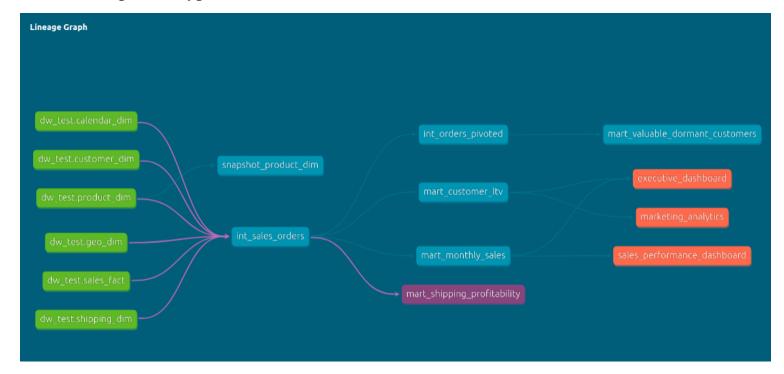
#### • Слой intermediate

Промежуточные модели не предназначены для конечных пользователей. Их цель — инкапсулировать сложную, многократно используемую бизнеслогику (например, JOIN'ы), чтобы не повторять код в каждой витрине

#### • Слой marts

Витрины данных предоставляют конечным пользователям быстрые и понятные данные для ВІ и аналитики. Представляют собой физические таблицы, агрегированные и оптимизированные для конкретных бизнес-задач.

### 2. Архитектура DWH



- 3. Ключевые фрагменты кода
- Код промежуточной модели int\_sales\_orders.sql

#### **SELECT**

-- Ключи

f.order\_id,

-- Измерения из customer\_dim

c.customer\_id,

c.customer\_name,

```
-- Измерения из product_dim
  p.product_id,
  p.product_name,
  p.category,
  p.sub_category,
  p.segment,
  -- Измерения из geo dim
  g.city,
  g.state,
  -- Измерения из shipping_dim
  s.ship_mode,
  -- Даты из calendar dim (с правильными псевдонимами)
  cal_order.date as order_date,
  cal_ship.date as ship_date,
  -- Метрики из sales_fact
  f.sales,
  f.profit,
  f.quantity,
  f.discount
FROM {{ source('dw_test', 'sales_fact') }} AS f
LEFT JOIN {{ source('dw_test', 'customer_dim') }} AS c ON f.cust_id = c.cust_id
LEFT JOIN {{ source('dw_test', 'product_dim') }} AS p ON f.prod_id = p.prod_id
LEFT JOIN {{ source('dw_test', 'shipping_dim') }} AS s ON f.ship_id = s.ship_id
LEFT JOIN {{ source('dw_test', 'geo_dim') }} AS g ON f.geo_id = g.geo_id
-- ИСПРАВЛЕНО: Добавляем псевдонимы, так как календарь используется дважды
LEFT JOIN {{ source('dw_test', 'calendar_dim') }} AS cal_order ON f.order_date_id =
cal order.dateid
LEFT JOIN {{ source('dw_test', 'calendar_dim') }} AS cal_ship ON f.ship_date_id =
cal_ship.dateid
```

### • Код индивидуальной mart-модели mart\_shipping\_profitability.sql

```
--Эффективность логистики. рассчитать общую прибыль для каждого
--способа доставки и сравнить ее со средней прибылью по компании.
WITH company_avg_profit AS (
  SELECT
    ROUND(AVG(profit), 2) AS avg_company_profit
  FROM {{ref('int_sales_orders')}}
),
shipping_profit AS (
  SELECT
    ship_mode,
    ROUND(SUM(profit), 2) AS total_profit,
    COUNT(DISTINCT order_id) AS order_count
  FROM {{ref('int_sales_orders')}}
  GROUP BY ship_mode
)
SELECT
  sp.ship_mode, --способ доставки
  sp.order_count, --количество заказов для способа доставки
  sp.total_profit, --общая прибыль способа доставки
  cap.avg_company_profit, --средняя прибыль по компании
  (sp.total_profit - cap.avg_company_profit) as profit_difference, --разница между общей
прибыли способа доставки и средней прибыли по компании
  CASE
    WHEN cap.avg_company_profit != 0 THEN
      ROUND(((sp.total_profit - cap.avg_company_profit) / cap.avg_company_profit) * 100, 2)
    ELSE 0
  END AS total_profit_percentage –процентное отклонение
FROM shipping_profit sp
CROSS JOIN company_avg_profit cap
ORDER BY sp.total_profit DESC
```

• Код кастомного теста test\_is\_positive.sql

```
-- tests/generic/test_is_positive.sql
{% test is_positive(model, column_name) %}
SELECT *
FROM {{ model }}
WHERE {{ column_name }} < 0
{% endtest %}</pre>
```

## • Код snapshot\_product\_dim.sql

```
-- snapshots/snapshot_product_dim.sql

{% snapshot snapshot_product_dim %}

{{
    config(
        target_schema='dw_snapshots',
        strategy='check',
        unique_key='prod_id',
        check_cols=['segment', 'category'],
    )

}}

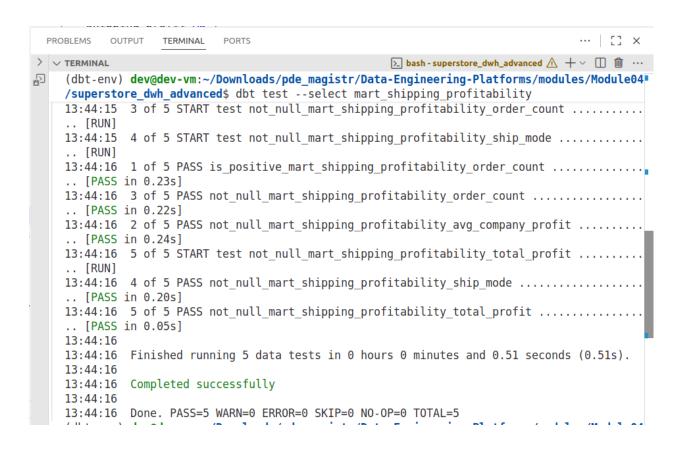
SELECT prod_id, product_id, segment, category FROM {{ source('dw_test', 'product_dim') }}

{% endsnapshot %}
```

# 4. Результаты

• Скриншоты успешного выполнения dbt run, dbt test, dbt snapshot

```
PROBLEMS
           OUTPUT
                  TERMINAL
                           PORTS
> V TERMINAL
                                                    (dbt-env) dev@dev-vm:~/Downloads/pde magistr/Data-Engineering-Platforms/modules/Module04
   /superstore dwh advanced$ dbt run --select mart shipping profitability
    13:43:24 Running with dbt=1.10.11
    13:43:24 Registered adapter: postgres=1.9.1
    13:43:25 [WARNING]: Configuration paths exist in your dbt project.yml file which do not
     apply to any resources.
    There are 1 unused configuration paths:
    - models.superstore dwh advanced.staging
    13:43:25 Found 6 models, 1 snapshot, 49 data tests, 6 sources, 3 exposures, 435 macros
    13:43:25
    13:43:25 Concurrency: 4 threads (target='dev')
    13:43:25
    13:43:25 1 of 1 START sql table model public dw test.mart shipping profitability ......
    .. [RUN]
    13:43:26  1 of 1 OK created sql table model public dw test.mart shipping profitability .
    .. [SELECT 4 in 1.02s]
    13:43:26
    13:43:26 Finished running 1 table model in 0 hours 0 minutes and 1.19 seconds (1.19s).
    13:43:26
    13:43:26 Completed successfully
    13:43:26
    13:43:26 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 NO-0P=0 TOTAL=1
```



```
(dbt-env) dev@dev-vm:~/Downloads/pde magistr/Data-Engineering-Platforms/modules/Module04
/superstore dwh advanced$ dbt snapshot
 13:45:38 Running with dbt=1.10.11
 13:45:38 Registered adapter: postgres=1.9.1
 13:45:39 [WARNING]: Configuration paths exist in your dbt project.yml file which do not
  apply to any resources.
 There are 1 unused configuration paths:
 models.superstore_dwh_advanced.staging
 13:45:39 Found 6 models, 1 snapshot, 49 data tests, 6 sources, 3 exposures, 435 macros
 13:45:39
 13:45:39 Concurrency: 4 threads (target='dev')
 13:45:39
 13:45:39 1 of 1 START snapshot dw snapshots.snapshot product dim ......
 .. [RUN]
 13:45:40 1 of 1 OK snapshotted dw snapshots.snapshot product dim ..........
 .. [INSERT 0 0 in 0.53s]
 13:45:40
 13:45:40 Finished running 1 snapshot in 0 hours 0 minutes and 0.71 seconds (0.71s).
 13:45:40
 13:45:40 Completed successfully
 13:45:40
 13:45:40 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=1
```

### • Скриншот с данными из индивидуальной mart-модели

<b>=</b> +		otifications				
	ship_mode character varying (14)	order_count bigint	total_profit numeric	avg_company_profit numeric	profit_difference numeric	total_profit_percentage numeric
1	Standard Class	3049	432025.65	29.36	431996.29	1471377.01
2	Second Class	1006	145980.38	29.36	145951.02	497108.38
3	First Class	842	132911.97	29.36	132882.61	452597.45
4	Same Day	312	48654.37	29.36	48625.01	165616.52

#### 5. Вывод

Использование промежуточных моделей и витрин позволяет создать многоуровневую архитектуру данных, которая абстрагирует сложную бизнеслогику, обеспечивает стандартизацию расчетов и значительно упрощает разработку аналитических запросов по сравнению с работой напрямую с единой таблицей фактов.