ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ И НАУКИ ГОРОДА МОСКВЫ
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»
(ГАОУ ВО МГПУ)
Институт цифрового образования
Департамент информатики, управления и технологий

Лабораторная работа № 2.1

по дисциплине «Платформы Data Engineering»

Выполнил:
студент группы БД251м

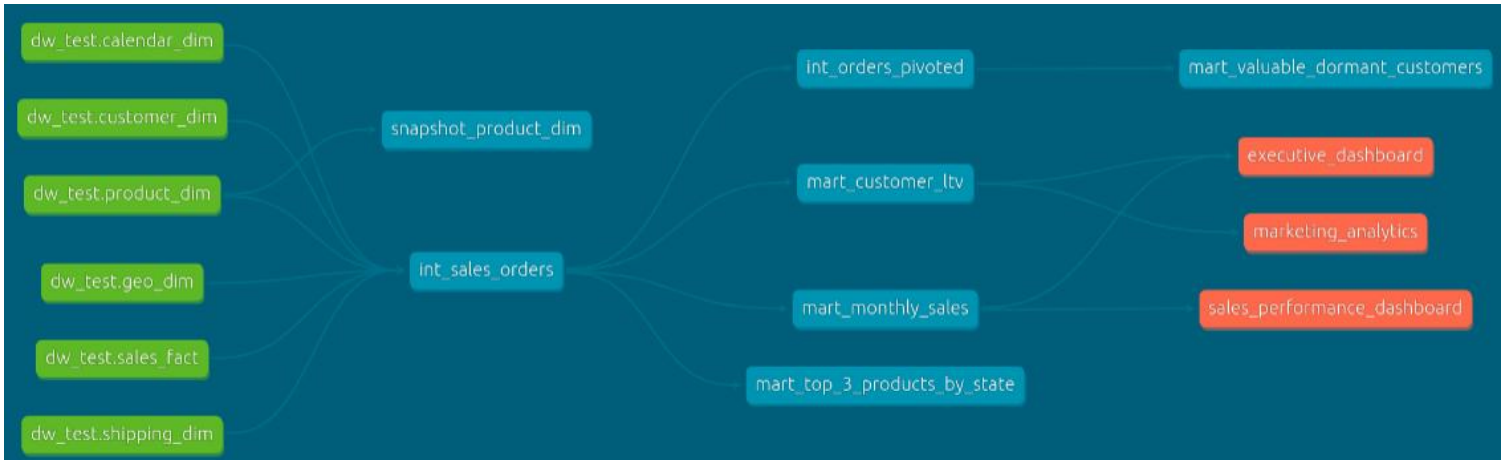
Направление подготовки/Специальность
38.04.05 - Бизнес-информатика
St_62
(Ф.И.О.)

Проверил:
Кандидат технических наук, доцент
(ученая степень, звание)
Босенко Тимур Муртазович
(Ф.И.О.)

Москва 2025

Архитектура dbt-проекта реализована по многоуровневому принципу: staging → intermediate → marts. Слой intermediate инкапсулирует сложную и многократно используемую бизнес-логику, а слой marts предоставляет бизнес-ориентированные витрины для BI и аналитики.

Архитектура DWH



Ключевые фрагменты кода

```
-- models/intermediate/int_sales_orders.sql
-- Эта модель объединяет факты со всеми измерениями, создавая
-- широкую, денормализованную таблицу для легкого использования в витринах.

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


-- Топ-3 самых прибыльных товаров по каждому штату.st_62

WITH product_profit AS (
SELECT
state,
product_id,
product_name,
SUM(profit) AS total_profit,
SUM(sales) AS total_sales
FROM {{ ref('int_sales_orders') }}
GROUP BY 1,2,3
),
ranked AS (
SELECT
state,
product_id,
product_name,
total_profit,
total_sales,
ROW_NUMBER() OVER (
PARTITION BY state
ORDER BY total_profit DESC, total_sales DESC, product_id
) AS rank_in_state
FROM product_profit
)
SELECT
state,
product_id,
product_name,
total_profit,
total_sales,
rank_in_state
FROM ranked
WHERE rank_in_state <= 3
ORDER BY state, rank_in_state



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

```yaml
  - name: mart_top_3_products_by_state
    description: "Топ-3 самых прибыльных товаров по каждому штату"
    columns:
      - name: state
        tests:
          - not_null
      - name: product_id
        tests:
          - not_null
      - name: product_name
        tests:
          - not_null
      - name: total_sales
        tests:
          - not_null
          - is_positive
```

```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 %}
```

Результаты

```
(dbt-env) dev@dev-vm:~/Downloads/pde_magistr/superstore_dwh_advanced$ dbt run
17:28:09  Running with dbt=1.10.11
17:28:09  Registered adapter: postgres=1.9.1
17:28:09  [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
17:28:09  Found 6 models, 1 snapshot, 52 data tests, 6 sources, 3 exposures, 435 macros
17:28:09
17:28:09  Concurrency: 4 threads (target='dev')
17:28:09
17:28:10  1 of 6 START sql view model public_dw_intermediate.int_sales_orders ............ [RUN]
17:28:10  1 of 6 OK created sql view model public_dw_intermediate.int_sales_orders ....... [CREATE VIEW in 0.11s]
17:28:10  2 of 6 START sql view model public_dw_intermediate.int_orders_pivoted .......... [RUN]
17:28:10  3 of 6 START sql table model public_dw_test.mart_customer_ltv .................. [RUN]
17:28:10  4 of 6 START sql table model public_dw_test.mart_monthly_sales .................. [RUN]
17:28:10  5 of 6 START sql table model public_dw_test.mart_top_3_products_by_state ....... [RUN]
17:28:10  2 of 6 OK created sql view model public_dw_intermediate.int_orders_pivoted ..... [CREATE VIEW in 0.21s]
17:28:10  6 of 6 START sql table model public_dw_test.mart_valuable_dormant_customers .... [RUN]
17:28:10  5 of 6 OK created sql table model public_dw_test.mart_top_3_products_by_state .. [SELECT 145 in 0.33s]
17:28:10  3 of 6 OK created sql table model public_dw_test.mart_customer_ltv ............. [SELECT 1193 in 0.49s]
17:28:10  4 of 6 OK created sql table model public_dw_test.mart_monthly_sales ............ [SELECT 688 in 0.46s]
17:28:10  6 of 6 OK created sql table model public_dw_test.mart_valuable_dormant_customers  [SELECT 299 in 0.33s]
17:28:10
17:28:10  Finished running 4 table models, 2 view models in 0 hours 0 minutes and 0.86 seconds (0.86s).
17:28:10
17:28:10  Completed successfully
17:28:10
17:28:10  Done. PASS=6 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=6
```

```
(dbt-env) dev@dev-vm:~/Downloads/pde_magistr/superstore_dwh_advanced$ dbt test
17:29:18  Running with dbt=1.10.11
17:29:18  Registered adapter: postgres=1.9.1
17:29:18  [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
17:29:19  Found 6 models, 1 snapshot, 52 data tests, 6 sources, 3 exposures, 435 macros
17:29:19
17:29:19  Concurrency: 4 threads (target='dev')
17:29:19
17:29:19  1 of 52 START test accepted_values_mart_top_3_products_by_state_rank_in_state__1__2__3  [RUN]
17:29:19  2 of 52 START test accepted_values_mart_valuable_dormant_customers_customer_status__DORMANT_VALUABLE  [RUN
]
17:29:19  3 of 52 START test is_positive_mart_customer_ltv_average_order_value ........... [RUN]
17:29:19  4 of 52 START test is_positive_mart_customer_ltv_number_of_orders .............. [RUN]
17:29:19  1 of 52 PASS accepted_values_mart_top_3_products_by_state_rank_in_state__1__2__3  [PASS in 0.13s]
17:29:19  3 of 52 PASS is_positive_mart_customer_ltv_average_order_value ................. [PASS in 0.14s]
17:29:19  5 of 52 START test is_positive_mart_customer_ltv_total_sales_lifetime .......... [RUN]
17:29:19  6 of 52 START test is_positive_mart_monthly_sales_number_of_orders ............. [RUN]
17:29:19  2 of 52 PASS accepted_values_mart_valuable_dormant_customers_customer_status__DORMANT_VALUABLE  [PASS in 0
.14s]
17:29:19  7 of 52 START test is_positive_mart_monthly_sales_total_sales .................. [RUN]
17:29:19  4 of 52 PASS is_positive_mart_customer_ltv_number_of_orders .................... [PASS in 0.18s]
17:29:19  8 of 52 START test is_positive_mart_top_3_products_by_state_total_sales ........ [RUN]
17:29:19  6 of 52 PASS is_positive_mart_monthly_sales_number_of_orders ................... [PASS in 0.14s]
17:29:19  5 of 52 PASS is_positive_mart_customer_ltv_total_sales_lifetime ................ [PASS in 0.16s]
17:29:19  9 of 52 START test is_positive_mart_valuable_dormant_customers_avg_order_value . [RUN]
17:29:19  10 of 52 START test is_positive_mart_valuable_dormant_customers_days_since_last_order  [RUN]
17:29:19  7 of 52 PASS is_positive_mart_monthly_sales_total_sales_....................... [PASS in 0.14s]
17:29:21  42 of 52 START test not_null_mart_valuable_dormant_customers_first_order_date .. [RUN]
17:29:21  40 of 52 START test not_null_mart_valuable_dormant_customers_customer_status ... [RUN]
17:29:21  39 of 52 PASS not_null_mart_valuable_dormant_customers_customer_name ........... [PASS in 0.11s]
17:29:21  43 of 52 START test not_null_mart_valuable_dormant_customers_last_order_date ... [RUN]
17:29:21  42 of 52 PASS not_null_mart_valuable_dormant_customers_first_order_date ........ [PASS in 0.11s]
17:29:21  44 of 52 START test not_null_mart_valuable_dormant_customers_sales_percentile .. [RUN]
17:29:21  41 of 52 PASS not_null_mart_valuable_dormant_customers_days_since_last_order ... [PASS in 0.14s]
17:29:21  40 of 52 PASS not_null_mart_valuable_dormant_customers_customer_status ......... [PASS in 0.12s]
17:29:21  45 of 52 START test not_null_mart_valuable_dormant_customers_segment ........... [RUN]
17:29:21  43 of 52 PASS not_null_mart_valuable_dormant_customers_last_order_date ......... [PASS in 0.11s]
17:29:21  46 of 52 START test not_null_mart_valuable_dormant_customers_state ............. [RUN]
17:29:21  47 of 52 START test not_null_mart_valuable_dormant_customers_total_orders ...... [RUN]
17:29:21  47 of 52 PASS not_null_mart_valuable_dormant_customers_total_orders ............ [PASS in 0.09s]
17:29:21  46 of 52 PASS not_null_mart_valuable_dormant_customers_state ................... [PASS in 0.10s]
17:29:21  48 of 52 START test not_null_mart_valuable_dormant_customers_total_profit ...... [RUN]
17:29:21  44 of 52 PASS not_null_mart_valuable_dormant_customers_sales_percentile ........ [PASS in 0.16s]
17:29:21  45 of 52 PASS not_null_mart_valuable_dormant_customers_segment ................. [PASS in 0.14s]
17:29:21  49 of 52 START test not_null_mart_valuable_dormant_customers_total_quantity .... [RUN]
17:29:21  51 of 52 START test unique_mart_customer_ltv_customer_id ....................... [RUN]
17:29:21  50 of 52 START test not_null_mart_valuable_dormant_customers_total_sales ....... [RUN]
17:29:21  48 of 52 PASS not_null_mart_valuable_dormant_customers_total_profit ............ [PASS in 0.15s]
17:29:21  49 of 52 PASS not_null_mart_valuable_dormant_customers_total_quantity .......... [PASS in 0.12s]
17:29:21  52 of 52 START test unique_mart_valuable_dormant_customers_customer_id ......... [RUN]
17:29:21  51 of 52 PASS unique_mart_customer_ltv_customer_id ............................. [PASS in 0.14s]
17:29:21  50 of 52 PASS not_null_mart_valuable_dormant_customers_total_sales ............. [PASS in 0.14s]
17:29:21  52 of 52 PASS unique_mart_valuable_dormant_customers_customer_id ............... [PASS in 0.07s]
17:29:21
17:29:21  Finished running 52 data tests in 0 hours 0 minutes and 2.38 seconds (2.38s).
17:29:21
17:29:21  Completed successfully
17:29:21
17:29:21  Done. PASS=52 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=52
```

```
(dbt-env) dev@dev-vm:~/Downloads/pde_magistr/superstore_dwh_advanced$ dbt snapshot
17:31:03  Running with dbt=1.10.11
17:31:03  Registered adapter: postgres=1.9.1
17:31:04  [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
17:31:04  Found 6 models, 1 snapshot, 52 data tests, 6 sources, 3 exposures, 435 macros
17:31:04
17:31:04  Concurrency: 4 threads (target='dev')
17:31:04
17:31:04  1 of 1 START snapshot dw_snapshots.snapshot_product_dim ..... [RUN]
17:31:04  1 of 1 OK snapshotted dw_snapshots.snapshot_product_dim ..... [INSERT 0 0 in 0.28s]
17:31:04
17:31:04  Finished running 1 snapshot in 0 hours 0 minutes and 0.48 seconds (0.48s).
17:31:04
17:31:04  Completed successfully
17:31:04
17:31:04  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=1
```

	state character varying (20)	product_id character varying (15)	product_name character varying (127)	total_profit numeric	total_sa numeric
1	Alabama	TEC-MA-10000010	Hewlett-Packard Deskjet 3050a All-in-One Color Inkjet Printer	1459.2000000000000000	3040
2	Alabama	TEC-AC-10003870	Logitech Z-906 Speaker sys - home theater - 5.1-CH	1055.9680000000000000	2639
3	Alabama	FUR-TA-10001086	SAFCO PlanMaster Boards, 60w x 37-1/2d, White Melamine	948.4176000000000000	3647
4	Arizona	TEC-PH-10002624	Samsung Galaxy S4 Mini	634.4865000000000000	5639
5	Arizona	FUR-CH-10004287	SAFCO Arco Folding Chair	447.4439999999997000	3977
6	Arizona	OFF-PA-10004359	Multicolor Computer Printout Paper	440.3700000000000000	1258
7	Arkansas	OFF-BI-10001359	GBC DocuBind TL300 Electric Binding System	1686.3412000000000000	3587
8	Arkansas	OFF-BI-10004318	Ibico EB-19 Dual Function Manual Binding System	636.6032000000000000	1383
9	Arkansas	TEC-PH-10002563	Adtran 1202752G1	458.6036000000000000	1763
10	California	OFF-BI-10003527	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	6863.3460000000000000	18302
11	California	TEC-CO-10001449	Hewlett Packard LaserJet 3310 Copier	6551.8908000000000000	18719
12	California	TEC-CO-10003763	Canon PC1060 Personal Laser Copier	4535.9352000000000000	13439

	product_id character varying (15)	product_name character varying (127)	total_profit numeric	total_sales numeric	rank_in_state bigint
1	TEC-MA-10000010	Hewlett-Packard Deskjet 3050a All-in-One Color Inkjet Printer	1459.2000000000000000	3040.0000	1
2	TEC-AC-10003870	Logitech Z-906 Speaker sys - home theater - 5.1-CH	1055.9680000000000000	2639.9200	2
3	FUR-TA-10001086	SAFCO PlanMaster Boards, 60w x 37-1/2d, White Melamine	948.4176000000000000	3647.7600	3
4	TEC-PH-10002624	Samsung Galaxy S4 Mini	634.4865000000000000	5639.8800	1
5	FUR-CH-10004287	SAFCO Arco Folding Chair	447.4439999999997000	3977.2800	2
6	OFF-PA-10004359	Multicolor Computer Printout Paper	440.3700000000000000	1258.2000	3
7	OFF-BI-10001359	GBC DocuBind TL300 Electric Binding System	1686.3412000000000000	3587.9600	1
8	OFF-BI-10004318	Ibico EB-19 Dual Function Manual Binding System	636.6032000000000000	1383.9200	2
9	TEC-PH-10002563	Adtran 1202752G1	458.6036000000000000	1763.8600	3
10	OFF-BI-10003527	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	6863.3460000000000000	18302.2560	1
11	TEC-CO-10001449	Hewlett Packard LaserJet 3310 Copier	6551.8908000000000000	18719.6880	2
12	TEC-CO-10003763	Canon PC1060 Personal Laser Copier	4535.9352000000000000	13439.8080	3

Выводы

Преимущество использования промежуточных моделей и витрин по сравнению с работой напрямую с единой таблицей фактов:

- Запросы в BI становятся короткими и понятными, снижается порог ошибок.
- Правка в одном месте автоматически распространяется на все витрины.
- Каждый человек может настроить витрину под свои требования.