

# Тестирование ETL перед выкатом в прод

- [Стандартный сценарий тестирования](#)
- [Сценарий для более длительного тестирования](#)
- [FAQ по тестированию](#)

## Стандартный сценарий тестирования

Чтобы протестировать даг необходимо:

- 1) Проверить что у вашей сервисной УЗ есть права на все таблицы к которым вы обращаетесь.
- 2) Прогнать даг в стейдж среде

Как это сделать, сейчас разберем. Представим, что вы разработали следующий dag:

```

... #imports

--  CH_ANALYTICS__BI__CONN_ID

#stage_table = stage.analytics__first_orders #
stage_table = sandbox_dev.stage__first_orders_tst #    sandbox_dev

#target_table = analytics.first_orders #
target_table = sandbox_dev.first_orders_tst #    sandbox_dev

dag = DAG(
    ...
)

sql = '''
INSERT INTO {target_table}
WITH payments AS (
    SELECT
        id AS payments_id,
        order_id,
        payment_method_id,
        source_id,
        source_type
    FROM analytics.int_spree_payments
)
SELECT
    max(payments_id) AS id,
    order_id,
    argMax(payment_method_id, payments_id) AS payment_method_id,
    argMax(source_id, payments_id) AS source_id,
    argMax(source_type, payments_id) AS source_type
FROM payments
GROUP BY order_id
'''

def write_data():
    # 1) stage
    ch_pd_read_sql_query(sql="TRUNCATE TABLE {stage_table}", ch_conn_id=CH_ANALYTICS__BI__CONN_ID)

    # 2) stage
    ch_pd_read_sql_query(sql=sql.format(target_table=stage_table), ch_conn_id=CH_ANALYTICS__BI__CONN_ID)

    # 3)
    ch_pd_read_sql_query(sql="ALTER TABLE {target_table} REPLACE PARTITION tuple() FROM {stage_table}",
ch_conn_id=CH_ANALYTICS__BI__CONN_ID)

    # 4) stage
    ch_pd_read_sql_query(sql="TRUNCATE TABLE {stage_table}", ch_conn_id=CH_ANALYTICS__BI__CONN_ID)

#    target_table = sandbox_dev.first_orders_tst
load_data_to_prod = PythonOperator(
    task_id="load_data_to_prod",
    python_callable=write_data,
    dag=dag
)

```

**Прежде чем его тестировать нужно убедиться что у продовой УЗ хватает прав на чтение и запись.**

В данном примере нужно убедиться, что у продовой УЗ есть права на запись в `analytics.first_orders`, `stage.analytics__first_orders` и на чтение в `analytics.int_spree_payments`

Если у УЗ есть права на запись в таблицу, то по умолчанию есть и на чтение этой же таблицы.

Сделать это можно так:

```
SELECT
    format('{0}.{1}', database, `table`) as table_name, access_type
FROM system.grants g
WHERE user_name='service.prod.analytics.bi.airflow-analytics'
    AND table_name in ('analytics.first_orders', 'analytics.int_spree_payments')
```

Имя своей сервисной УЗ можно найти в [реестре сервисных УЗ](#) по имени коннекшена в аирфлоу.

### **Важно!!!**

#### **В коннектах stage airflow прописаны креды от stage УЗ**

Для каждой prod УЗ в airflow существует stage УЗ (в реестре можно увидеть что они находятся рядом)

сервисная stage УЗ умеет читать все таблицы, которые может читать и писать прод УЗ

сервисная stage УЗ НЕ умеет писать в прод таблицы, только в [sandbox\\_dev](#) , [sandbox\\_research](#)

Если у УЗ не хватает прав на чтение или запись, то нужно запросить права через dwh-support.

### **Когда убедились, что прав хватает, тестируем ETL в sandbox\_dev**

В sandbox\_dev создаем таблицу для записи [sandbox\\_dev.first\\_orders\\_tst](#)  
И прогоняем даг на stage airflow

## Сценарий для более длительного тестирования

Допустим если вам нужно собрать витрину, и показать данные витрины за несколько дней, то тестируем в схеме [sandbox\\_research](#), то есть пишем данные в таблицу в этой схеме

**Обратите внимание на именование таблиц в этой схеме!** Полное описание в документе по ссылке. Обратите внимание на именование таблиц в этой схеме!

## FAQ по тестированию

### Q: Почему не могу писать в таблицу в stage схеме из stage airflow?

А: Таблицы в stage схеме являются частью прод процессов, представьте, у вас есть прод процесс который пишет промежуточные данные в stage а потом переносит их в dm схему ЕСЛИ в это же самое время в stage airflow будет работать процесс который использует ту же таблицу, то как минимум мы можем потерять данные в stage а в худшем случае зальем дубли или не корректные данные в dm схему.

### Q: Как тогда тестировать даг в stage airflow?

А: Создавать аналог таблицы в которую нужно писать в схеме [sandbox\\_dev](#) / [sandbox\\_research](#) .

### Q: Как тестировать даг с ПД таблицами, если stage учета не может их читать?

А: Создайте полную копию таблицы в sandbox\_dev с хешами вместо колонок с ПД данными.