Документация.

1. Слой STG.

Слой stg состоит из трех таблиц БД DE в Postgresql: stg.api\_restaurants, stg.api\_couriers, stg.api\_deliveries.

Данные в них попадают из API курьерской службы соответствующими GET методами: /restaurants, /couriers, /deliveries

Посредством оркестратора AirFLow на языке python. Dag содержит такси\задачи на каждую загрузку в таблицу.

Данные заливаются As Is без каких-либо изменений.

Использованы следующие библиотеки python: requests, psycopg2, PostgresHook

1. Слой DDS.

Слой состоит также из трех соответствующих таблиц: dm\_api\_restaurants, dm\_api\_couriers, dm\_api\_deliveries.

В первых двух таблицах реализован SCD2 подход с перезаписью в случае прихода того же самого id ресторана или id курьера

С изменением полей active\_from (проставляется now() апдейта таблицы) и active\_to (проставляется 2099-01-01)

В таблице dm\_api\_deliveries происходит перезапись заказов (по проверке уникальности order\_id) за последние 7 дней начиная с уже залитой максимальной даты.

1. Слой CDM.

В этом слое собирается финальная витрина, которая содержит всю необходимую информацию.

Делается проверка на уникальность связки courier\_id + settlement\_year + settlement\_month и перезаливаются агрегированные данные.