

6. OLTP→OLAP

Автоматизация

Я разделил импорт на достаточно большое количество файлов, чтобы в проекте было проще ориентироваться. Ничего страшного в их количестве нету, каждый из них просто отвечает за свою часть процесса. Предполагается, что файлы будут запускаться все вместе, желательно одной командой.

Речь про набор файлов в этой папке:

```
./etl/oltpToOlap
```

В случае, если вам неудобно запускать такое большое количество файлов - проще всего будет немного модифицировать один из скриптов командной строки, который я уже написал. Эти два скрипта (под Windows и Mac/Linux) делают одно и то же: выполняют на сервере с конкретными credentials все скрипты в конкретной директории в порядке их очередности.

Найти оба скрипта можно тут, с названиями "fireOltpToOlap" и соответствующими расширениями:

```
./cmd/fireOltpToOlap.sh  
./cmd/fireOltpToOlap.bat
```

Они изначально предназначены для Docker, поэтому потребуется заменить credentials базы и путь к папке со скриптами в вашей OS, а также убрать начальную часть с docker exec, вызывающую команду внутри конкретного контейнера.

P.S. Скрипты нужно запускать на OLAP-базе, не на OLTP!

Про сами скрипты ETL

Примечательного тут немного. Я использую Foreign Data Wrapper для доступа к серверу базы oltp_db и создаю Foreign Tables для доступа к данным, а затем наполняю схему. Внешние таблицы, с которыми идет работа в каждом скрипте - инициализируются заново в начале скрипта во избежание проблем.

Что требует внимания: первый скрипт - 01_init_fdw.sql содержит параметры для подключения к базе данных OLTP. В вашем случае необходимо использовать credentials от сервера oltp_db.

```
CREATE SERVER oltp_server  
  FOREIGN DATA WRAPPER postgres_fdw  
  OPTIONS (host 'oltp', port '5432', dbname 'oltp_db');  
-- Replace 'oltp' with localhost  
-- Replace 'port' with oltp server port  
-- Replace 'dbname' if you give oltp database another name
```

```
CREATE USER MAPPING FOR CURRENT_USER  
  SERVER oltp_server  
  OPTIONS (user 'oltp_user', password 'oltp_password');  
-- Replace credentials with oltp user creds, if changed
```

Также этот скрипт проверяет доступ к серверу OLTP и сообщает, если что-то пошло не так. Последующие скрипты используют уже настроенное подключение и содержат только бизнес-логику

В некоторых случаях скрипты работают в несколько шагов, особенно это касается импортов в таблицы с SCD Type 2 - так просто более понятно, когда мы сначала закрываем старые записи, а потом назначаем новые. Я также постарался оптимизировать процесс, где это возможно - например, не запрашивая старые записи, если это не требуется для контекста.

И наверное ещё пару слов про даты. 3-й скрипт содержит заполнение dim_date на 5 лет вперед относительно текущей даты и потенциально может быть исключен из большинства запусков ETL, потому что одного заполнения вполне достаточно. Он также оптимизирован под заполнение только необходимых дат.

В качестве вывода

По окончании всех скриптов получаем базу данных OLAP, готовую к работе для аналитиков. Если что-то пошло не так - первое, что стоит проверить:

- Ошибку подключения к OLTP - она есть?
- Credentials, если ошибка присутствует.