



Зеркалирование в Clickhouse и создание витрины данных

Цель: получить данные из PostgreSQL в Clickhouse и построить витрину данных, представляющую собой набор агрегаций по месяцам

Описание задания:

В предыдущей задаче вы создали таблицу `customs_log` и витрину данных по ней `monthly_trade_summary` в Postgres. Теперь задача состоит в том, чтобы найти способ зеркалировать данные из `customs_log` в одноименную и такую же по содержанию таблицу в Clickhouse. Соединение может производиться например с помощью функции [postgresql](#) или [табличного движка](#).

Затем в Clickhouse создается витрина данных `monthly_trade_summary` с агрегированным отчетом по месяцам

month: месяц в исходном формате из лога (yyyy/mm)

total_import_value: общий объем импорта в единицах (см. примечание ниже) за этот месяц

total_export_value: общий объем экспорта в единицах (см. примечание ниже) за этот месяц

trade_balance: разница между `total_export_value` и `total_import_value`

top_import_category: категория, товары из которой чаще всего импортировались в этом месяце

top_export_category: категория, товары из которой чаще всего экспортировались в этом месяце

major_import_partner: код страны, откуда было больше всего импорта в этом месяце

major_export_partner: код страны, куда было больше всего экспорта в этом месяце

total_transactions: общее количество операций экспорта и импорта за этот месяц

ПРИМЕЧАНИЯ: для упрощения задачи можно игнорировать поле `measure` и использовать только поле `quantity`, считая все в единицах

Можно разделить действия создания и вставки данных - создать в Clickhouse таблицу на движке MergeTree и вставить в неё данные из PostgreSQL.

На свое усмотрение добавляете индексирование и MATERIALIZED VIEW (особенно если в предыдущий раз вы выгрузили не часть данных, а весь датасет, рекомендуется оптимизировать работу с данными).

Ответом на задание является ссылка на Git-репозиторий, где лежит файл `monthly_trade_summary.txt` с кодом всех использованных в задании запросов, т.е.:

- зеркалирования таблицы `customs_log` (одним запросом или разделенными по действиям)
- (опционально) создание дополнительных индексов на таблице `customs_log`
- (опционально) создание MATERIALIZED VIEW для агрегации данных для расчета витрины
- создание таблицы `monthly_trade_summary`
- один запрос на заполнение таблицы `monthly_trade_summary` на основе таблицы `customs_log`

Инструменты, которые пригодятся для выполнения: Clickhouse, PostgreSQL.



Критерии, по которым будет оцениваться задание: имена таблиц, код запроса зеркалирования customs_log, а также финальный SQL-запрос должен генерировать корректную витрину в Clickhouse. Данные в таблицах customs_log в PostgreSQL и Clickhouse должны совпадать. Если при подготовке витрины были созданы индексы и materialized view для финального запроса файл должен содержать комментарии около соответствующих запросов.

Ваше решение

Отправить на проверку

Сохранить как черновик

Далее >>

Слёрм ©

[+7 \(495\) 248-05-80](tel:+7(495)248-05-80)

[Лицензия №ДЛ-1368 от 22.08.2019](#)

[Политика конфиденциальности](#)

[Публичная оферта](#)

