**Лабораторная работа №8**

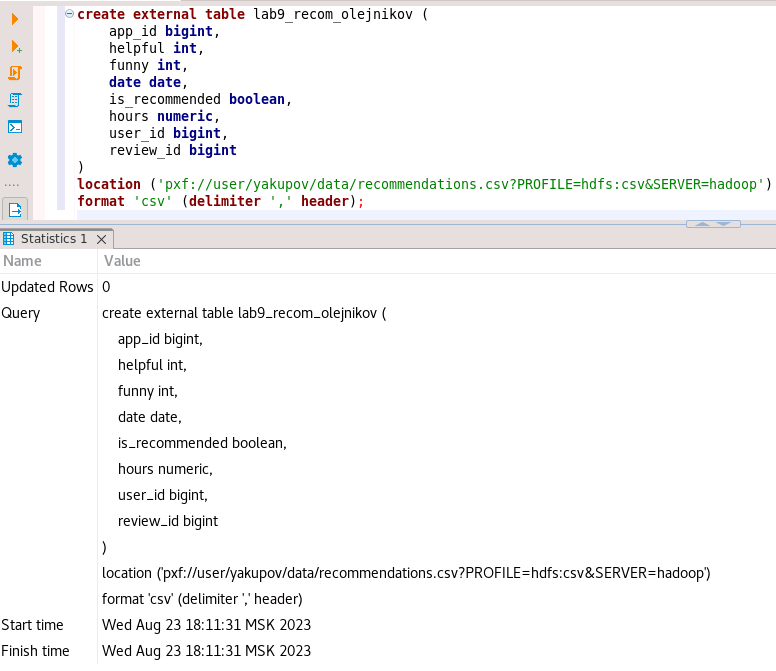
**Олейников Михаил Николаевич (olejnikov)**

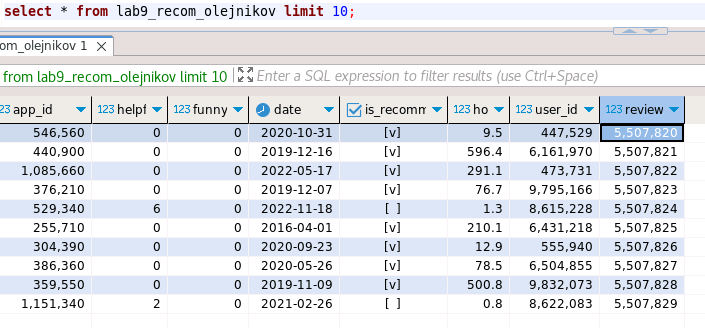
В текущей работе вам потребуется использовать продукт RT.WareHouse, который развернут в рамках нашего курса. Если появятся вопросы, пожалуйста напишите в общий учебный чат или личным сообщением преподавателю

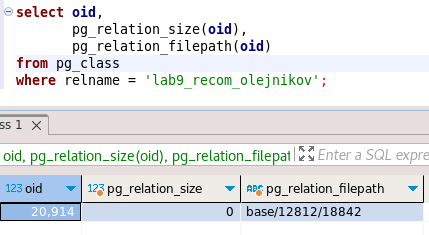
**Цель задания**: попрактиковаться с таблицами в GreenPlum и со схемой виртуализацией данных.

1. Используя HDFS файл **/user/yakupov/data/recommendations.csv** постройте интеграцию Greenplum -> HDFS, используя PXF подход и EXTERNAL таблицу.

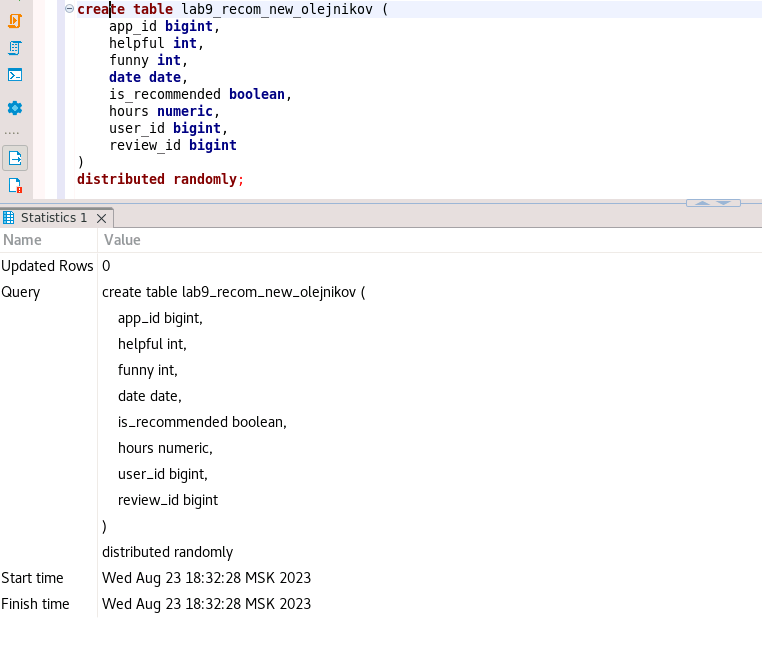
Назовите вашу EXTERNAL таблицу, используя шаблон именования “lab9\_recom\_фамилия”(например **lab9\_recom\_yakupov**)



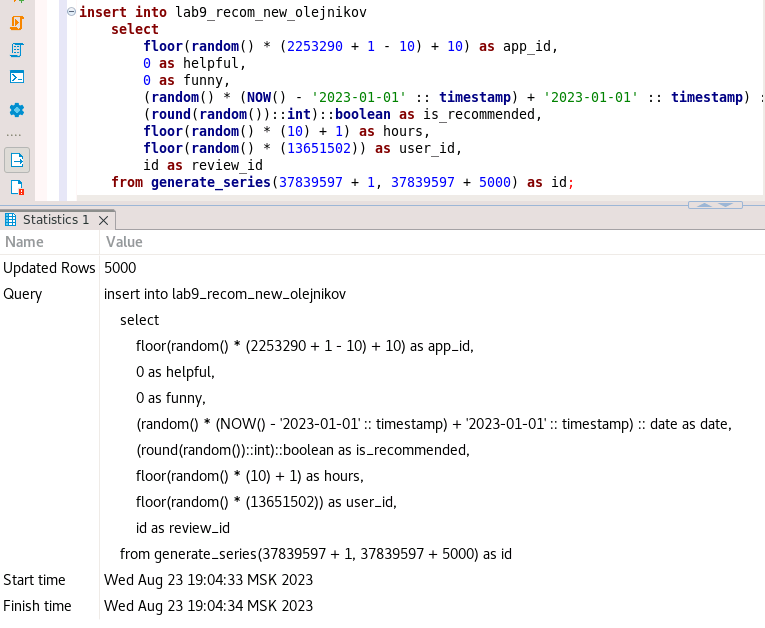


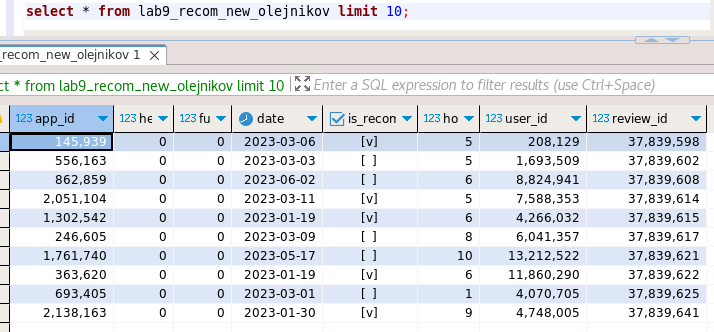


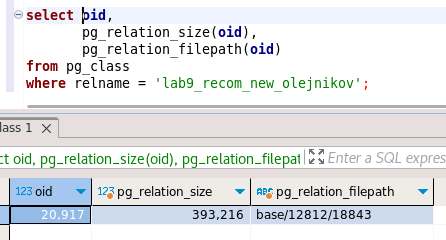
1. Создайте физическую таблицу в GreenPlum, используя шаблон именования “lab9\_recom\_new\_фамилия” (например **lab9\_recom\_new\_yakupov**) с моделью данных и типами абсолютно такими же как и у “lab9\_recom\_фамилия”. Выберите ключ дистрибуции, какой по вашему мнению необходим для предотвращения SKEW аномалии (самый простой - DISTRIBUTED RANDOMLY).



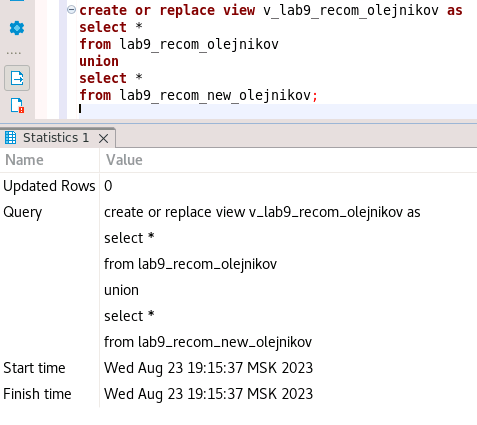
Заполните вашу физическую таблицу “lab9\_recom\_new\_фамилия” 5000 новыми строками, НО со значением поля модели данных “date” за 2023 год (чтобы данные из предоставленного файла и вновь сгенерированные не пересекались)

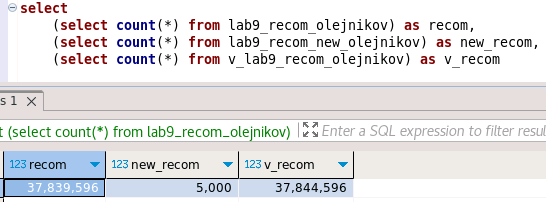






1. Создайте виртуальную таблицу VIEW с шаблоном именования “v\_lab9\_recom\_фамилия” (например **v\_lab9\_recom\_yakupov**), которая будет содержать UNION операцию для двух таблиц и пунктов a) и b)





(Предоставить преподавателю) приложите, пожалуйста, скрипты всех шагов в гугл документ и прикрепите ссылку в анкете. Не забудьте открыть доступ по ссылке.