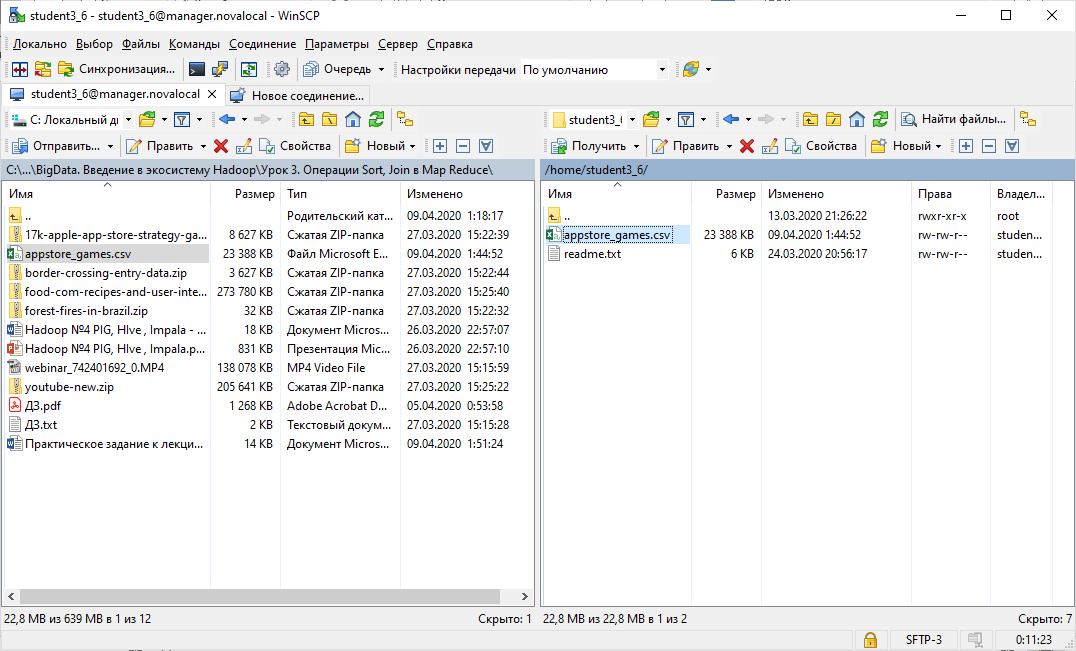
Практическое задание к лекции по HDFS (Лекция №3)

1. Скачать любой датасет из списка

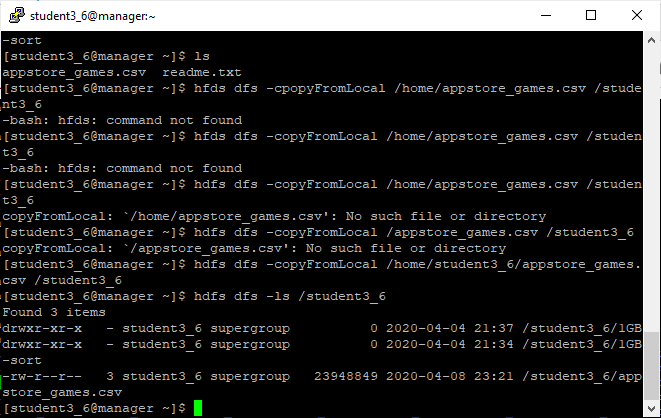
Скачал датасет: <https://www.kaggle.com/tristan581/17k-apple-app-store-strategy-games>

1. Загрузить этот датасет в HDFS в свою домашнюю папку

Загрузил на сервер с помощью WinSCP:

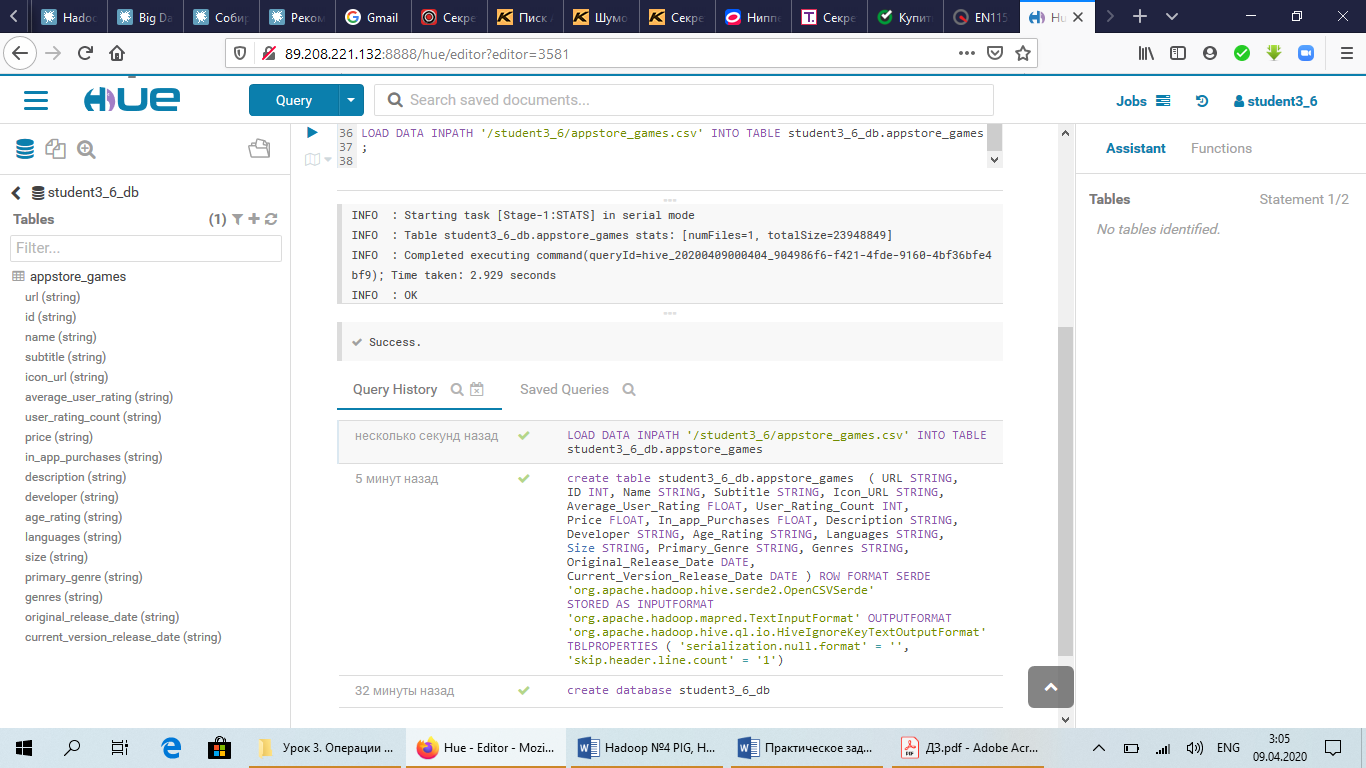


Потом передаю в hdfs:



1. Создать собственную базу данных в HIVE. Желательно, чтобы имя базы содержало номер вашего пользователя.

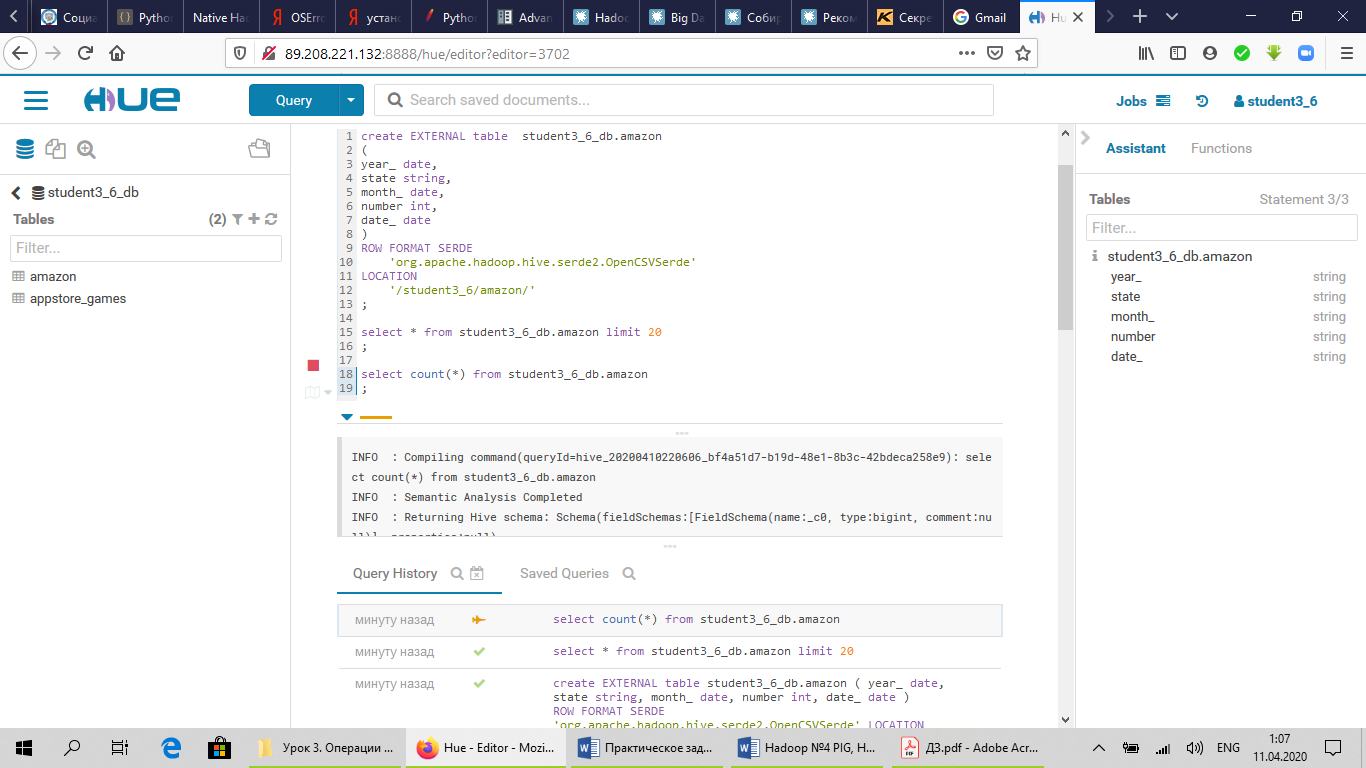
Создаём и наполняем базу данных на основе kaggle датасета 17k-apple-app-store-strategy-games:



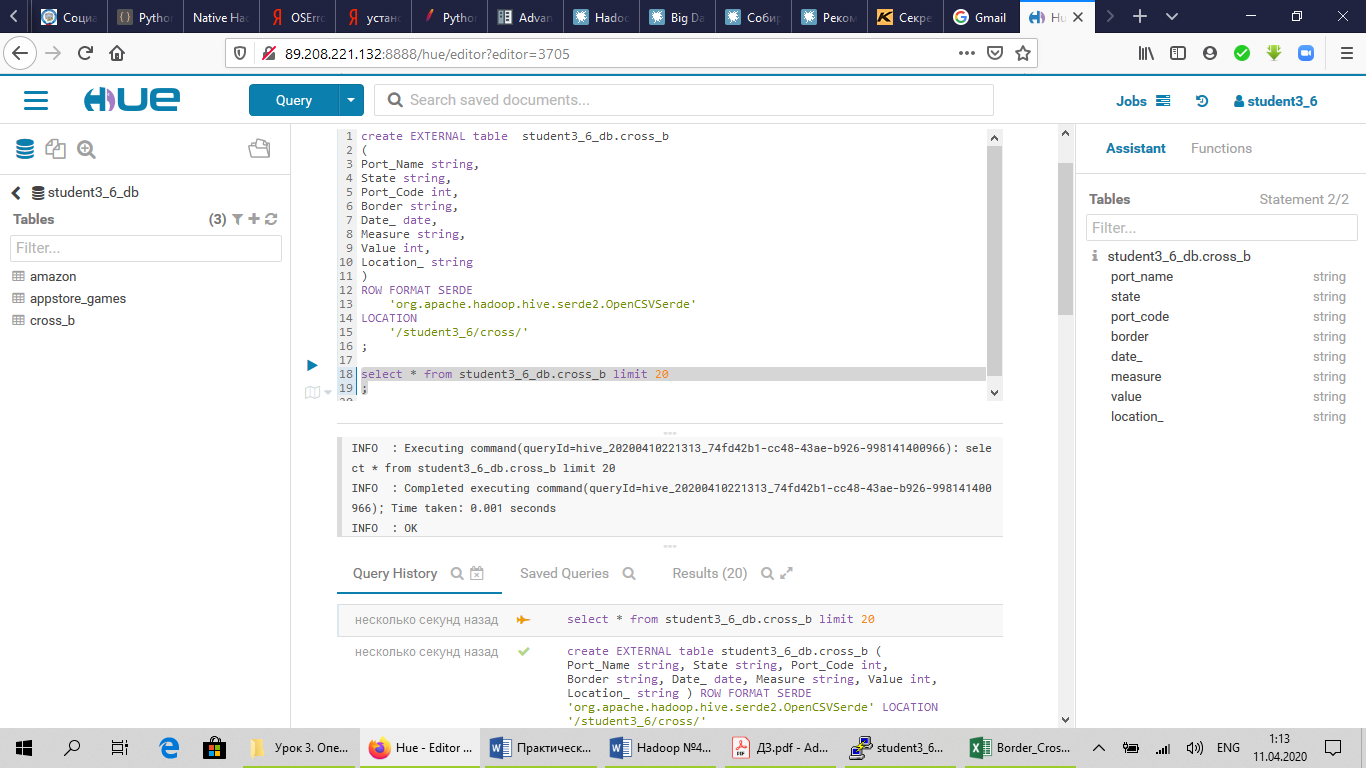
1. Создать EXTERNAL таблицы внутри базы данных с использованием всех загруженных файлов. Один файл – одна таблица.

На сервер подгрузил два датасета border-crossing-entry-data и forest-fires-in-brazil:

Первая таблица forest-fires-in-brazil

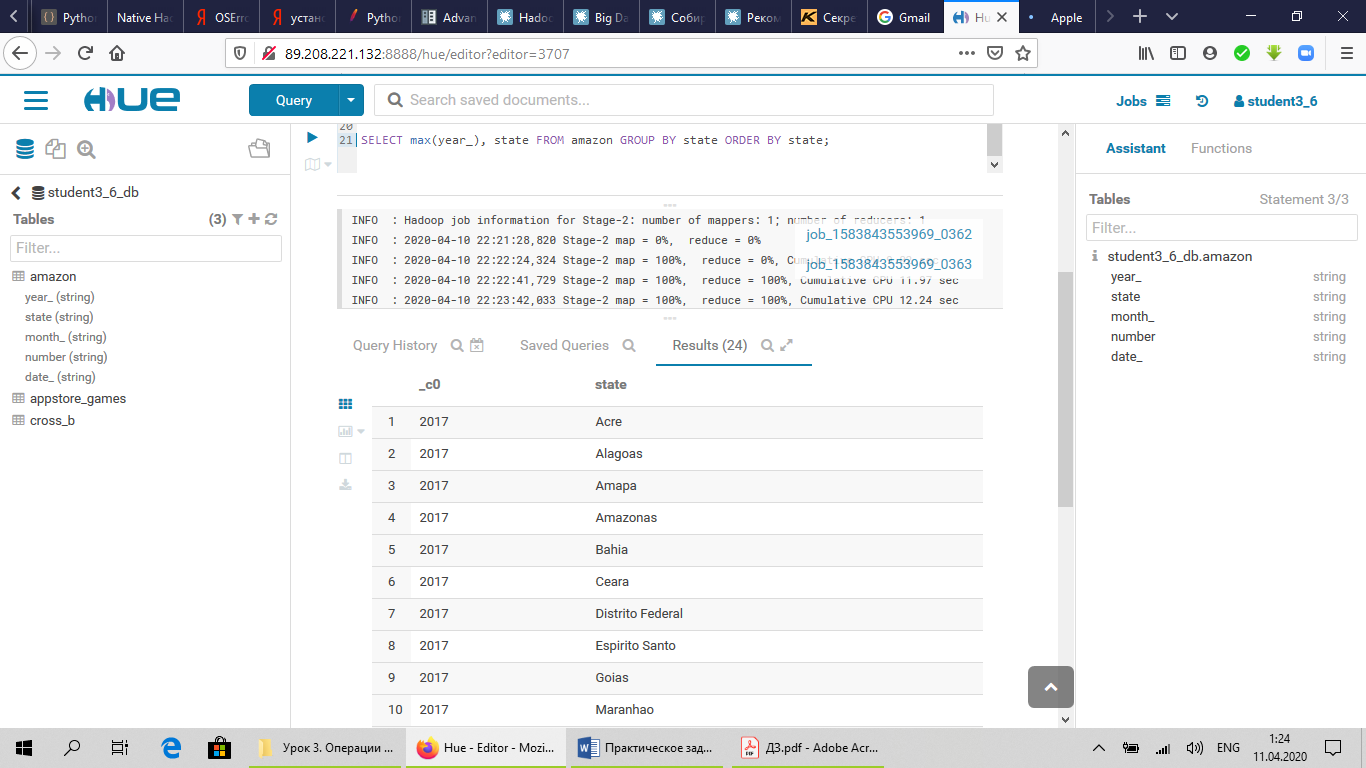


Вторая таблица border-crossing-entry-data

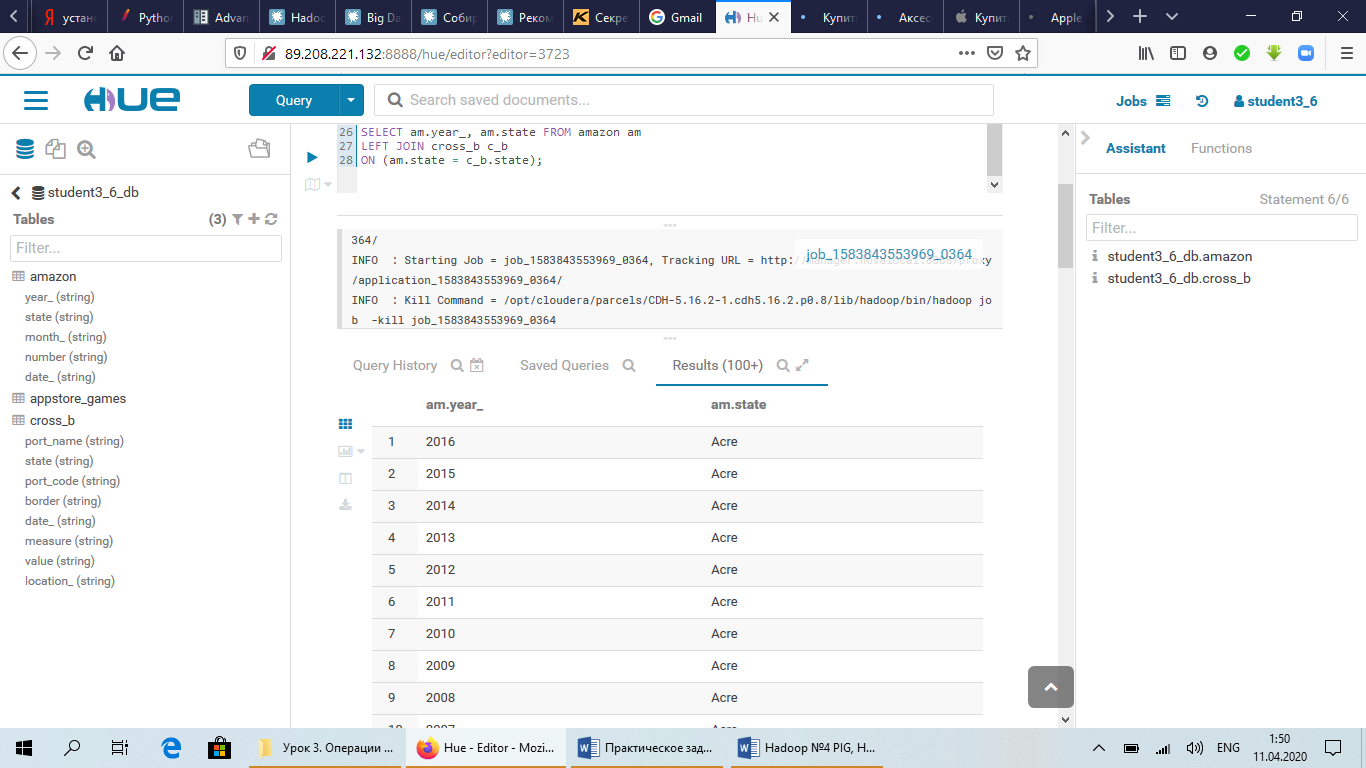


1. Сделать любой отчет по загруженным данным используя групповые и агрегатные функции.

Использовал в запросе сразу обе функции:



1. Сделать любой отчет по загруженным данным используя JOIN.



[Продвинутый вариант] Сделать все вышеперечисленное с использованием JSON SerDe. Подсказка: см в сторону команды «ADD JAR»