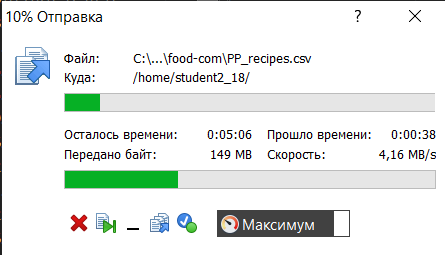
Скачал датасеты и загрузил их в «локалку» через WinScp

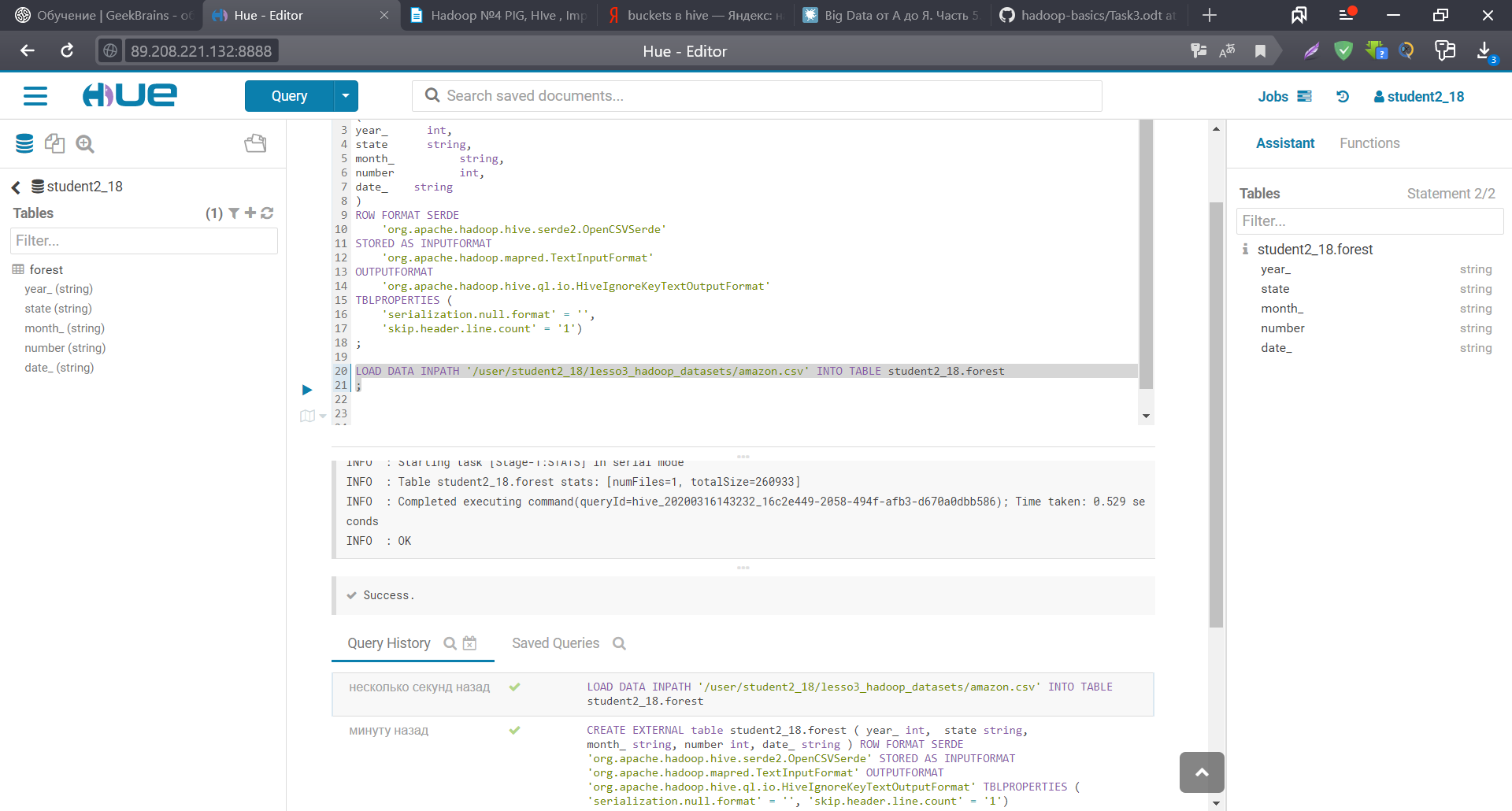


Скопировал их в HDFS



Создал БД, загрузил туда таблицы из скачанных датасетов:

forest



appstore

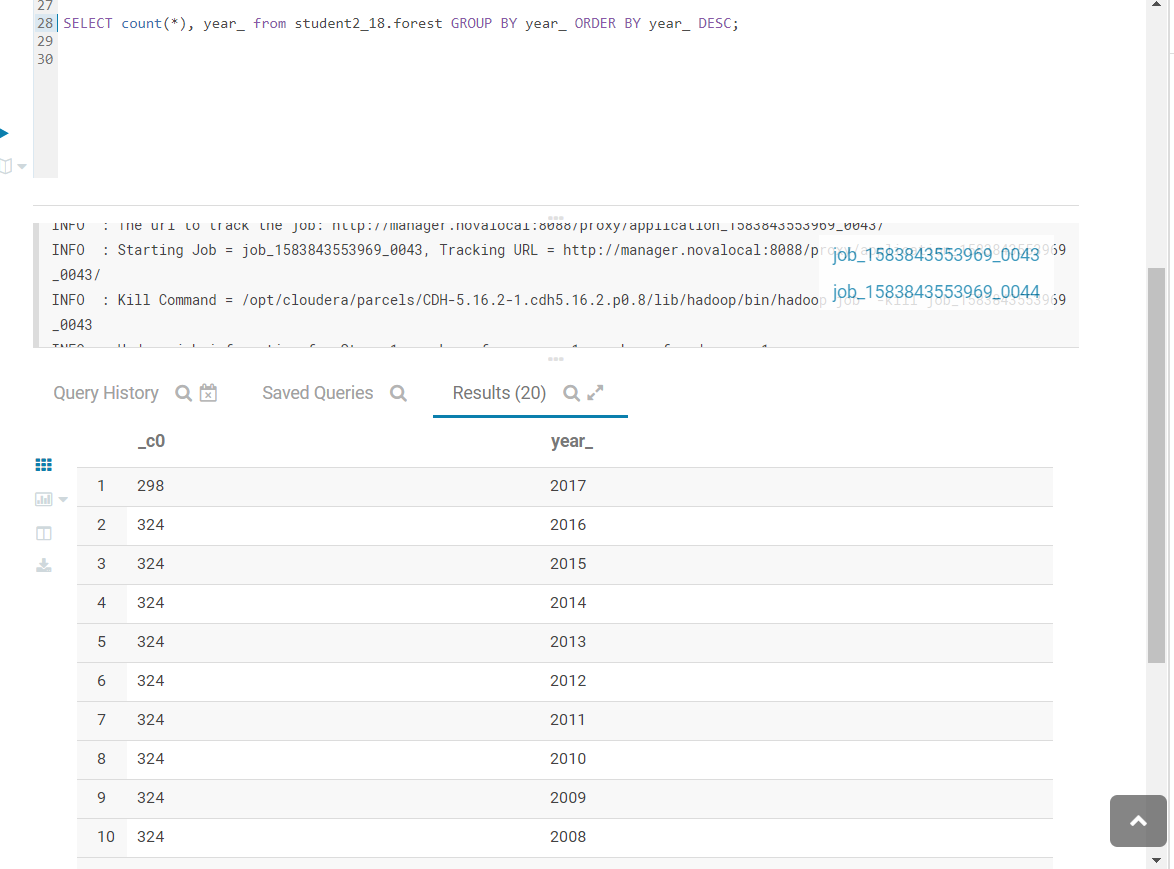


Border



Датасеты youtube и food не стал загружать, очень уж много табличек ;)

Отчет с использованием агрегационных функций



Не очень понял, как сделать отсчет, используя join, когда нет связанных между собой таблиц, но сделал самообъединение таблицы

