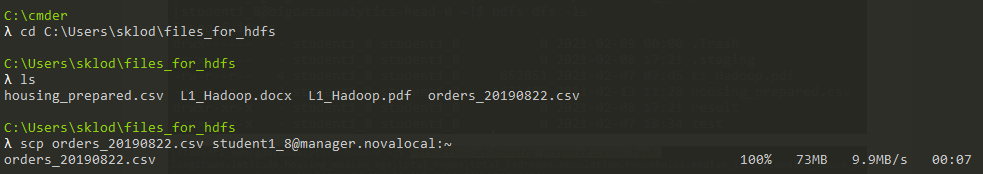
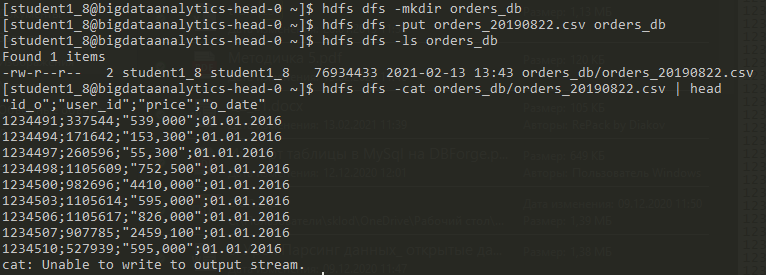
***Скачать любой датасет из списка ниже с сайта Kaggle.com (достаточно большой)***

*Я взяла датасет. Который мы использовали ранее на курсе аналитики баз данных* 

***Загрузить этот датасет в HDFS в свою домашнюю папку***

***Создала папку для датасета сразу и загрузила датасет в нее***



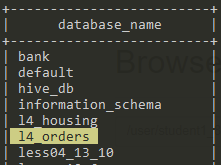
***Создать собственную базу данных в HIVE***

*Не буду скринить, т.к. на каждую строку кода выводится очень много текста)*

**[student1\_8@bigdataanalytics-head-0 ~]$ hive**

**0: jdbc:hive2://bigdataanalytics-worker-0.nov> create database l4\_orders;**

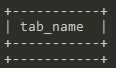
**0: jdbc:hive2://bigdataanalytics-worker-0.nov> show databases;**



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

**0: jdbc:hive2://bigdataanalytics-worker-0.nov> use l4\_orders;**

**0: jdbc:hive2://bigdataanalytics-worker-0.nov> show tables;**



*Переписываю код для создания таблицы исходя из своего датасета*

drop table if exists **l4\_orders.test\_orders;**

create external table **l4\_orders.test\_orders**

**(id\_o string,**

**user\_id string,**

**price string,**

**o\_date string)**

ROW FORMAT serde 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'

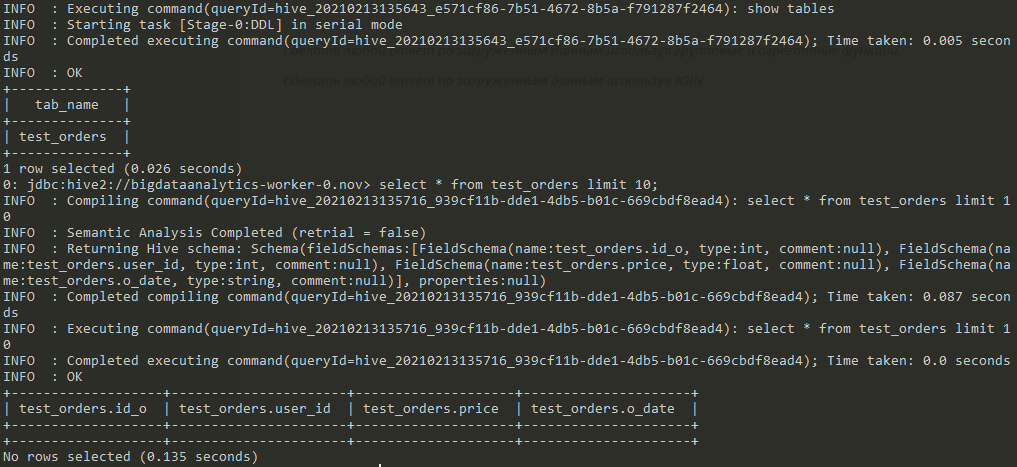
WITH SERDEPROPERTIES ('field.delim' = '**;**')

location "**/user/student1\_8/orders\_db**"

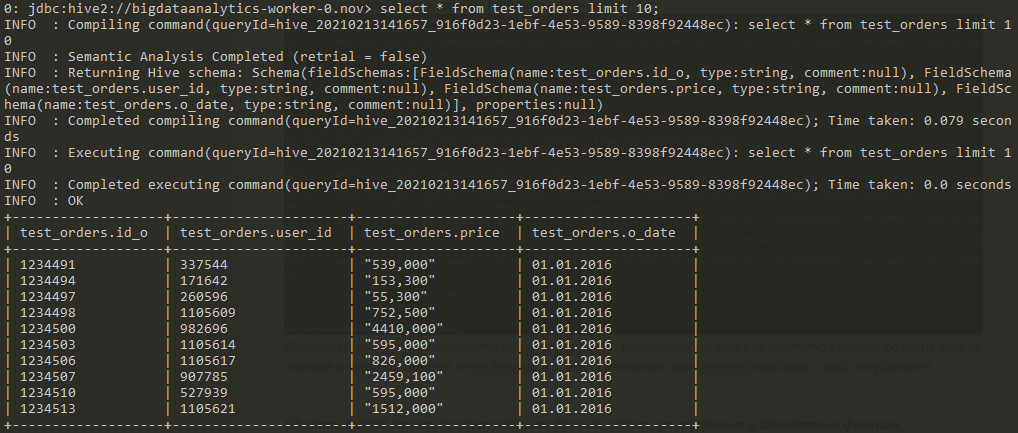
tblproperties ("skip.header.line.count"="**1**");

***я потом еще экспериментировала и в другой базе данных для одной из колонок хотела сделать тип данных BIT, так в ней хранятся только значения 1 и 0, но постоянно возникала ошибка – поменяла на INT и все починилось. Почему тип BIT не воспринимает?***

*Итак, методом проб и ошибок таблица наконец создана*



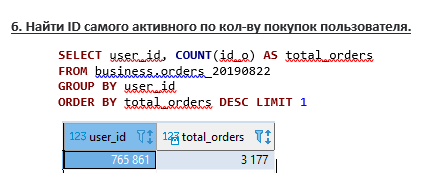
***Данные из датасета почему-то не подгрузились…*** *разобралась и еще раз наглядно увидела разницу между manage и external. Данные не подгрузились, т.к. я несколько раз дропала таблицу и мой csv удалился.*



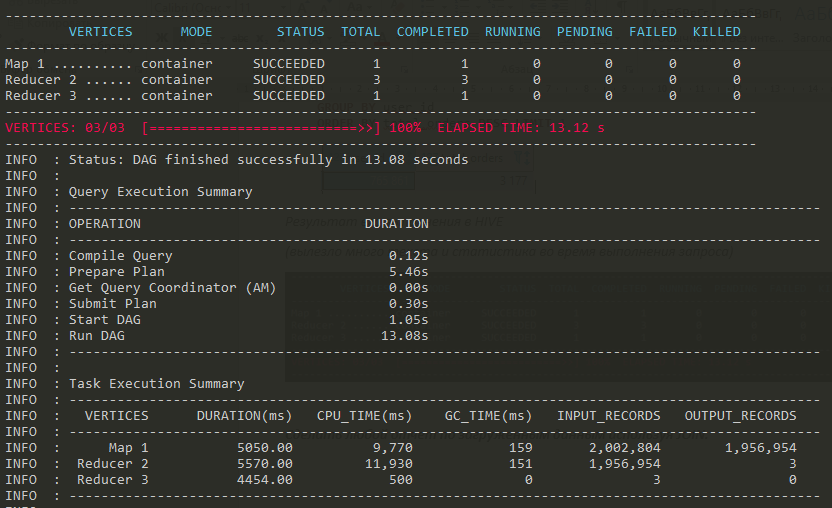
***Сделать любой отчет по загруженным данным используя групповые и агрегатные функции.***

*Поскольку SQL запросы не самоцель в данном ДЗ, я использовала уже имеющиеся у меня задачи с решениями к данному датасету*

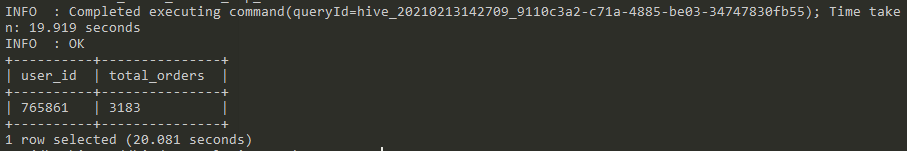
*Запрос, который у меня был*



*(вылезло много текста и статистика во время выполнения запроса)*



*Результат его выполнения в HIVE*



*Круто! Немного разнится кол-во заказов, но это норм, т.к. ранее я делала предобработку датасета и убирала заказы на сумму менее 100, а сейчас использовала необработанный датасет.*

***Сделать любой отчет по загруженным данным используя JOIN.***

*Я побоялась использовать пример который у меня был, т.к. там были представления и выборки по дате, а в таблице, которую я создала в HIVE я указала типа данных для дат строчный. Новую выборку с JOIN я не успела придумать*