Hive

1. Сначала скачал **docker-hive-master**
   удалил старый контейнер
   docker kill $(docker ps -q)

```
max@max:~/Downloads$ docker kill $(docker ps -q)
c089ebc74652
83a0c93b70db
897babfbfa9d
d8815c3996c1
5d686970546f
7a0c188d2378
max@max:~/Downloads$
```

   Зайду в папку, содержащую **hive** и запущу контейнер
   cd docker-hive-master

   поднял контейнер
   docker-compose up -d

```
max@max:~/Downloads/docker-hive-master$ docker-compose up -d
Starting docker-hive-master_datanode_1                ... done
Starting docker-hive-master_hive-server_1             ... done
Starting docker-hive-master_hive-metastore-postgresql_1 ... done
Starting docker-hive-master_presto-coordinator_1      ... done
Starting docker-hive-master_namenode_1                ... done
Starting docker-hive-master_hive-metastore_1          ... done
max@max:~/Downloads/docker-hive-master$
```

   Подключаюсь к неймноде
   docker exec -it docker-hive-master_namenode_1 bash

```
max@max:~/Downloads/docker-hive-master$ docker exec -it docker-hive-master_namen
ode_1 bash
root@5d686970546f:/#
```

2. Копирую файо АЭРОПОРТЫ в контейнер
   docker cp "/home/max/Downloads/airports.csv" 5d686970546f:/

```
max@max:~/Downloads$ docker cp "/home/max/Downloads/airports.csv" 5d686970546f:/
```

   Проверю

```
root@5d686970546f:/# ls
airports.csv  dev           hadoop       lib    mnt   root    sbin  tmp
bin           entrypoint.sh hadoop-data  lib64  opt   run     srv   usr
boot          etc           home         media  proc  run.sh  sys   var
root@5d686970546f:/#
```

3. Можем просмотреть, что записано в файле аэопорт
   head -5 airports.csv

4. Кладем АЭОПОРТ.csv на HDFS:
   hdfs dfs -put airports.csv /

```
root@5d686970546f:/# hdfs dfs -put airports.csv /
root@5d686970546f:/#
```

5. Создал папку «my_airports», перенес туда файл «airports.csv»
   hdfs dfs –mkdir /my_airports
   hdfs dfs -mv /airports.csv /my_airports

6. Открыл ДБивер и законектился

1. Подключиться к серверу, ввести команду hive, по очереди выполнить все команды из HSQL_1.sql, заменив student41_35 на своего пользователя. Сказать, какие действия мы выполняем в каждом из запросов/команд (на insert не забываем про tablesample). + ответить на вопросы, содержащиеся в HSQL_1.sql.

Ответ в файле HSQL_answers.sql

2. Сказать какие особенности, касающиеся синтаксиса HiveQL надо учитывать, создавая таблицу с партиционированием и при инсерте в неё. Зачем нам нужно партиционирование?

А) При создании убрать из create те поля, по которым будет партиционирование.

```
create table max_airports.airport_codes_part
   id int,
   ident string,
```

Б) Прописать поля по которым партиционируем в partitioned

```
partitioned by (`type` string) -- указали колонку, по которой будет партиционирование
stored as TEXTFILE
location '/hive_test_loc'      -- указал папку, в которой будут лежать данные для
                                  партиционирования
```

В) Партиционирование требуется для ускорения работы селектов (те данные, по которым будет большинство запросов включаем в партиц.)

3. Переписать «select …» в команде «create temporary table» используя “with” для объявления t2; пример:

```
Hive WITH clause example with the SELECT statement
WITH t1 as (SELECT 1),
t2 as (SELECT 2),
t3 as (SELECT 3)
SELECT * from t1
UNION ALL
SELECT * from t2
UNION ALL
SELECT * from t3;
```

*Создать таблицу airport_codes_part_2 с партиционированием по 2 колонкам: type и iso_country. Вставить в неё 1000 строк, запросом select вывести самую популярную связку type iso_country используя оконную функцию row_number https://www.revisitclass.com/hadoop/how-to-use-row-number-function-in-hive/.

Ответ в HSQL_4.sql