

Захожу в HIVE смотрю список БД

До этого момента раздал права на папку hdfs dfs -chmod 777 used_cars и файл hdfs dfs -chmod 777 used_cars/vehicles.csv

```
[student3_16@bigdataanalytics2-head-shdp1-v31-1-0 ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.1.4.0-315/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://bigdataanalytics2-worker-shdp1-v31-1-0.novalocal:2181,bigdataanalytics2-worker-shdp1-v31-1-3.novalocal:2181,bigdataanalytics2-worker-shdp1-v31-1-2.novalocal:2181,bigdataanalytics2-worker-shdp1-v31-1-1.novalocal:2181/default;password=zooKeeperNamespace=hiveserver2
21/04/19 15:06:27 [main]: INFO jdbc.HiveConnection: Connected to bigdataanalytics2-head-shdp1-v31-1-0.novalocal:10000
Connected to: Apache Hive (version 3.1.0.3.1.4.0-315)
Driver: Hive JDBC (version 3.1.0.3.1.4.0-315)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.1.4.0-315 by Apache Hive
0: jdbc:hive2://bigdataanalytics2-worker-shdp> show databases;
INFO : Compiling command(queryId=hive_20210419150641_62992959-8b02-4bdb-a895-747b9cf1d2c8): show databases
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20210419150641_62992959-8b02-4bdb-a895-747b9cf1d2c8); Time taken: 0.014 seconds
INFO : Executing command(queryId=hive_20210419150641_62992959-8b02-4bdb-a895-747b9cf1d2c8): show databases
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210419150641_62992959-8b02-4bdb-a895-747b9cf1d2c8); Time taken: 0.017 seconds
INFO : OK
+-----+
| database_name |
+-----+
| 2_18          |
| antong        |
| arture_less4  |
| avito_dataset |
| books         |
| chicago_crimes|
| climb        |
| dataexplorer  |
| db_student21_19|
| default       |
| dmitriyt      |
| dubcow        |
| dubcow1       |
| dubcow4       |
| fdc           |
| games         |
| happiness     |
| hive_db       |
| homework      |
| igord         |
| information_schema|
| job           |
| kaggle        |
| lelikov_a     |
| lesson3       |
| lesson3_kirillr|
| lesson3_sergeys|
| lesson4       |
| lesson4_kirillr|
+-----+
```

Создаю БД student3_16

```
0: jdbc:hive2://bigdataanalytics2-worker-shdp> create database student3_16
+-----+
| student3_16 |
+-----+
INFO : Compiling command(queryId=hive_20210419150839_00e90de7-381a-4473-8727-0ff83bf6c358): create database student3_16
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20210419150839_00e90de7-381a-4473-8727-0ff83bf6c358); Time taken: 0.029 seconds
INFO : Executing command(queryId=hive_20210419150839_00e90de7-381a-4473-8727-0ff83bf6c358): create database student3_16
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210419150839_00e90de7-381a-4473-8727-0ff83bf6c358); Time taken: 0.062 seconds
INFO : OK
No rows affected (0.135 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> █
```

```
+-----+
| student3_16 |
| student3_23 |
| student3_26 |
| student3_39 |
| student3_4   |
| student3_8   |
| sys          |
| testdb       |
| titanic      |
| trending_youtube_videos|
| vaccin_db    |
| valerias_hive_db|
| vorozhbitov  |
| vyacheslavk2 |
| wine         |
| zzz          |
+-----+
90 rows selected (0.054 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> █
```

Выбираю БД student3_16 и смотрю список доступных таблиц в БД

```
90 rows selected (0.054 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> use student3_16
INFO : Compiling command(queryId=hive_20210419150959_773c0e56-f9eb-4fa4-8d6b-d703ac7db7be): use student3_16
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20210419150959_773c0e56-f9eb-4fa4-8d6b-d703ac7db7be); Time taken: 0.008 seconds
INFO : Executing command(queryId=hive_20210419150959_773c0e56-f9eb-4fa4-8d6b-d703ac7db7be): use student3_16
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210419150959_773c0e56-f9eb-4fa4-8d6b-d703ac7db7be); Time taken: 0.01 seconds
INFO : OK
No rows affected (0.048 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> show tables;
INFO : Compiling command(queryId=hive_20210419151007_e460a6d4-cc7d-40f4-8825-73cb0f4d3824): show tables
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20210419151007_e460a6d4-cc7d-40f4-8825-73cb0f4d3824); Time taken: 0.008 seconds
INFO : Executing command(queryId=hive_20210419151007_e460a6d4-cc7d-40f4-8825-73cb0f4d3824): show tables
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210419151007_e460a6d4-cc7d-40f4-8825-73cb0f4d3824); Time taken: 0.006 seconds
INFO : OK
+-----+
| tab_name |
+-----+
+-----+
No rows selected (0.035 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> █
```

Создаю таблицу по заготовке:

```
create external table vehicles_2 (
```

```
  r_id string,
```

```
  id int,
```

```
  url string,
```

```
  region string,
```

```
  region_url string,
```

```
  price int,
```

```
  year date,
```

```
  manufacturer string,
```

```
  model string,
```

```
  condition string,
```

```
  cylinders int,
```

```
  fuel string,
```

```
  odometer float,
```

```
  title_status string,
```

```
  transmission string,
```

```
  VIN string,
```

```
  drive string,
```

```
  size string,
```

type string,

paint_color string,

image_url string,

description string,

state string,

lat string,

long string,

posting_date date

)

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

WITH SERDEPROPERTIES ('field.delim' = ',')

location "/user/student3_16/used_cars"

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

```
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20210419152156_11d76484-0873-46ad-9d99-0ca00fcd8adf); Time taken: 0.036 seconds
INFO : Executing command(queryId=hive_20210419152156_11d76484-0873-46ad-9d99-0ca00fcd8adf): create external table vehicles (
r_id string,
id int,
url string,
region string,
region_url string,
price int,
year date,
manufacturer string,
model string,
condition string,
cylinders int,
fuel string,
odometer float,
title_status string,
transmission string,
VIN string,
drive string,
size string,
type string,
paint_color string,
image_url string,
description string,
state string,
lat string,
long string,
posting_date date
)
partitioned by ( p_date string )
ROW FORMAT serde 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'
WITH SERDEPROPERTIES ('field.delim' = ',')
location "/user/student3_16/"
tblproperties ("skip.header.line.count"="1")
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210419152156_11d76484-0873-46ad-9d99-0ca00fcd8adf); Time taken: 0.106 seconds
INFO : OK
No rows affected (0.218 seconds)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> █
```

	vehicles_2.fuel	vehicles_2.model	vehicles_2.condition	vehicles_2.cylinders	vehicles_2.fuel	vehicles_2.odometer	vehicles_2.title_status	vehicles_2.transmission	vehicles_2.vin	vehicles_2.manufacture
	vehicles_2.drive	vehicles_2.size	vehicles_2.type	vehicles_2.paint_color	vehicles_2.image_url	vehicles_2.long	vehicles_2.description	vehicles_2.posting_date		vehicles_2
0	NULL	chevrolet	corvette grand sport	good	gas	32742.0	clean			other
1	1G1WJ30W4250090	rdw		good	NULL		https://images.craigslist.org/0080N/ikbWvZfzFg0c000x450.jpg	600x450.jpg	"Carvana is the safe	
er way to buy a car during these uncertain times Carvana is dedicated to ensuring safety for all of our customers. In addition to our 100% online shopping and selling experience that allows all customers									to buy and trade their cars without ever leaving the safety of their house we're providing touchless delivery that make all aspects of our process even safer. Now you can get the car you want	
1	NULL	hyundai	sonata	excellent	NULL	93600.0	clean	automatic		NULL
and try to sell my car. The car is in very good condition everything works and fully cleaned. It equipped with a heated seat power seat power seat backup camera							https://images.craigslist.org/0080s_gBHvN507vM0e00x450.jpg	600x450.jpg	"I'll move to another city	
2	NULL	x3 3.0i	good	NULL	gas	87046.0	clean	automatic		bm
rare Blue Water Metallic exterior and tan interior color combination. 5-speed automatic transmission AWD							https://images.craigslist.org/0080B_5z6G0P0r0_07l0k_600x450.jpg	"Clean 2006 BMW X3 3.0i. Beautiful and		CD/AM/FM radio
cold A/C (just serviced w/oil change)										chevrolet
3	NULL	c-10	full-size	good	pickap	106000.0	clean	automatic		NULL
t and back breaks al							https://images.craigslist.org/0080M_6o7K0P0Ac0_0c1r0_600x450.jpg	1974 chev. truck (LONG BED) NEW starter front		NULL
4	NULL	f350 larjat	full-size	pickap	blue	116000.0	clean	automatic		NULL
Proofed). This truck was bullet proofed early on and has been well maintained. Truck is equipped with a 6.6 liter turbo diesel. Currently has 116K miles. Everything on the truck works as it should truck										000 lbs
5	NULL	tacoma double cab sr5	pickap	red	good	32290.0	clean	other		NULL
6	3TM2425NGM020355	rdw		pickap	red	16922.0	https://images.craigslist.org/0080k_gnUNSLCBMT_0g0c0_600x450.jpg	"Carvana is the safer		
way to buy a car during these uncertain times Carvana is dedicated to ensuring safety for all of our customers. In addition to our 100% online shopping and selling experience that allows all customers to								to buy and trade their cars without ever leaving the safety of their house we're providing touchless delivery that make all aspects of our process even safer. Now you can get the car you want		
7	NULL	mustang shelby gt500 coupe	coupe	blue	NULL	9692.0	https://images.craigslist.org/01515_f1GMU8aZa0_0g0c0_600x450.jpg	"Carvana is the safer		
way to buy a car during these uncertain times Carvana is dedicated to ensuring safety for all of our customers. In addition to our 100% online shopping and selling experience that allows all customers to								to buy and trade their cars without ever leaving the safety of their house we're providing touchless delivery that make all aspects of our process even safer. Now you can get the car you want		

[illegible]

```
SELECT COUNT(manufacturer) FROM vehicles_2 WHERE manufacturer="tesla";
```

не дождался пока выполнится запрос, уж слишком долго выполняется

```

0: jdbc:hive2://bigdataanalytics2-worker-shdp> select manufacturer, sum(price) from vehicles_2 group by manufacturer order by manufacturer;
INFO : Compiling command(queryId=hive_20210419161148_6f6a3279-b86d-4e9e-b34f-36660162a0c6): select manufacturer, sum(price) from vehicles_2 group by manufacturer order by manufacturer
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:manufacturer, type:string, comment:null), FieldSchema(name:_c1, type:bigint, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20210419161148_6f6a3279-b86d-4e9e-b34f-36660162a0c6); Time taken: 0.155 seconds
INFO : Executing command(queryId=hive_20210419161148_6f6a3279-b86d-4e9e-b34f-36660162a0c6): select manufacturer, sum(price) from vehicles_2 group by manufacturer order by manufacturer
INFO : Query ID = hive_20210419161148_6f6a3279-b86d-4e9e-b34f-36660162a0c6
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20210419161148_6f6a3279-b86d-4e9e-b34f-36660162a0c6
INFO : Session is already open
INFO : Dag name: select manufacturer, sum(pric...manufacturer (Stage-1)

```

также не дождался пока выполнится запрос, уж слишком долго выполняется

Попробовал написать запрос с JOIN

SELECT vehicles_2.model, vehicles.year FROM vehicles_2 JOIN vehicles WHERE vehicles_2.manufacturer = vehicles.manufacturer;

```

Interrupting... Please be patient this may take some time.
Error: Query was cancelled (state=01000,code=0)
0: jdbc:hive2://bigdataanalytics2-worker-shdp> use student3_16
> ;
Interrupting... Please be patient this may take some time.
Interrupting... Please be patient this may take some time.

```

В итоге все окончательно повисло((