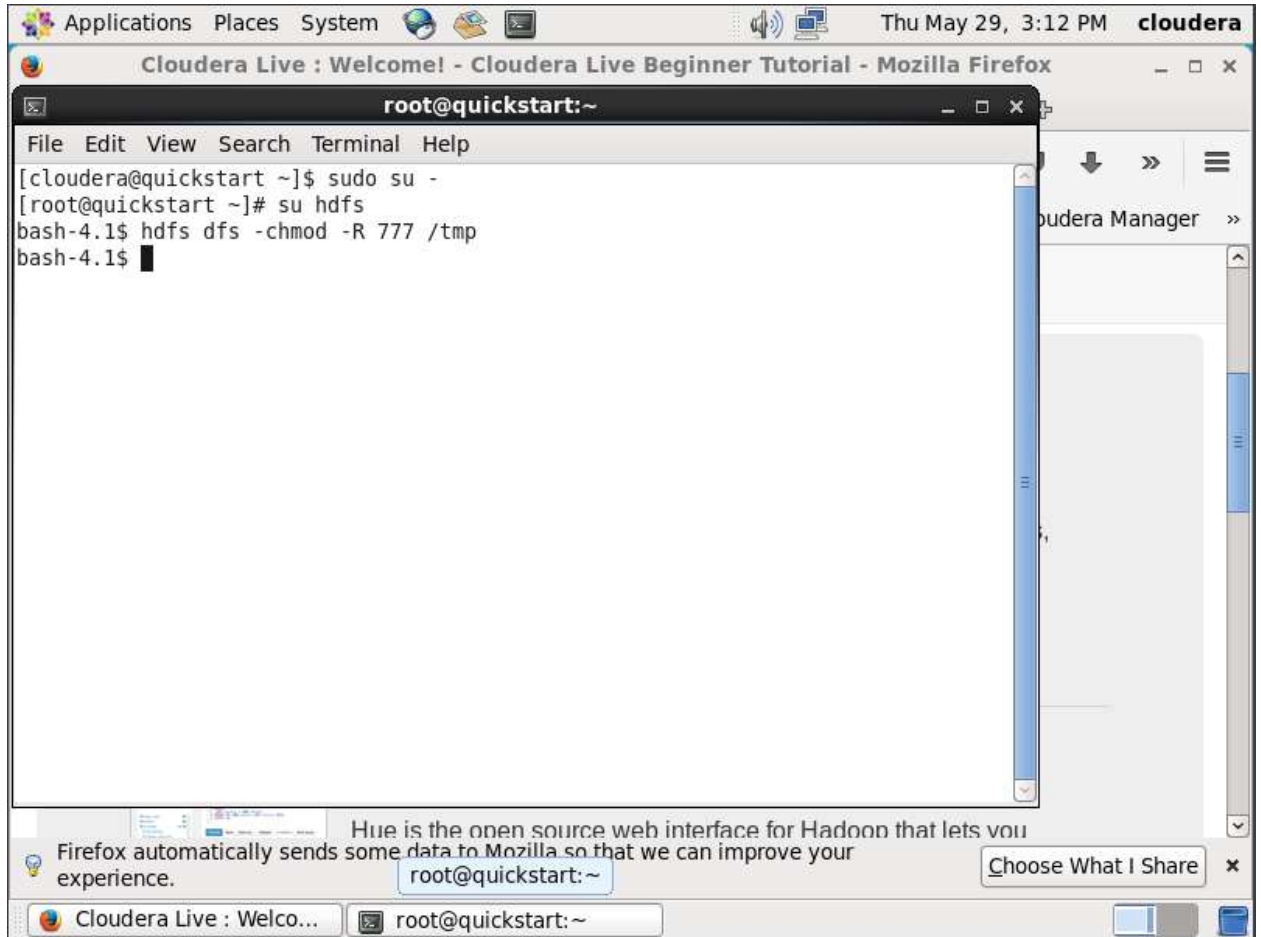


Устанавливаем разрешения через терминал

```
sudo su -
```

```
su hdfs
```

```
hdfs dfs -chmod -R 777 /tmp
```



Идем в браузер и нажимаем на установку hue

Applications Places System Thu May 29, 3:13 PM cloudera

Cloudera Live : Welcome! - Cloudera Live Beginner Tutorial - Mozilla Firefox

Cloudera Live : Welc... x Hue - Welcome t... x Cloudera Live : Welc... x Hue - Welcome t... x

quickstart.cloudera/#/ Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager

cloudera LIVE Navigation

Analyze Your Data

Hue is the open source web interface for Hadoop that lets you analyze your data. Simply load in your data and then easily begin to analyze, search, and visualize it. In the QuickStart VM, the administrative username for Hue is 'cloudera' and the password is 'cloudera'.

Launch Hue UI

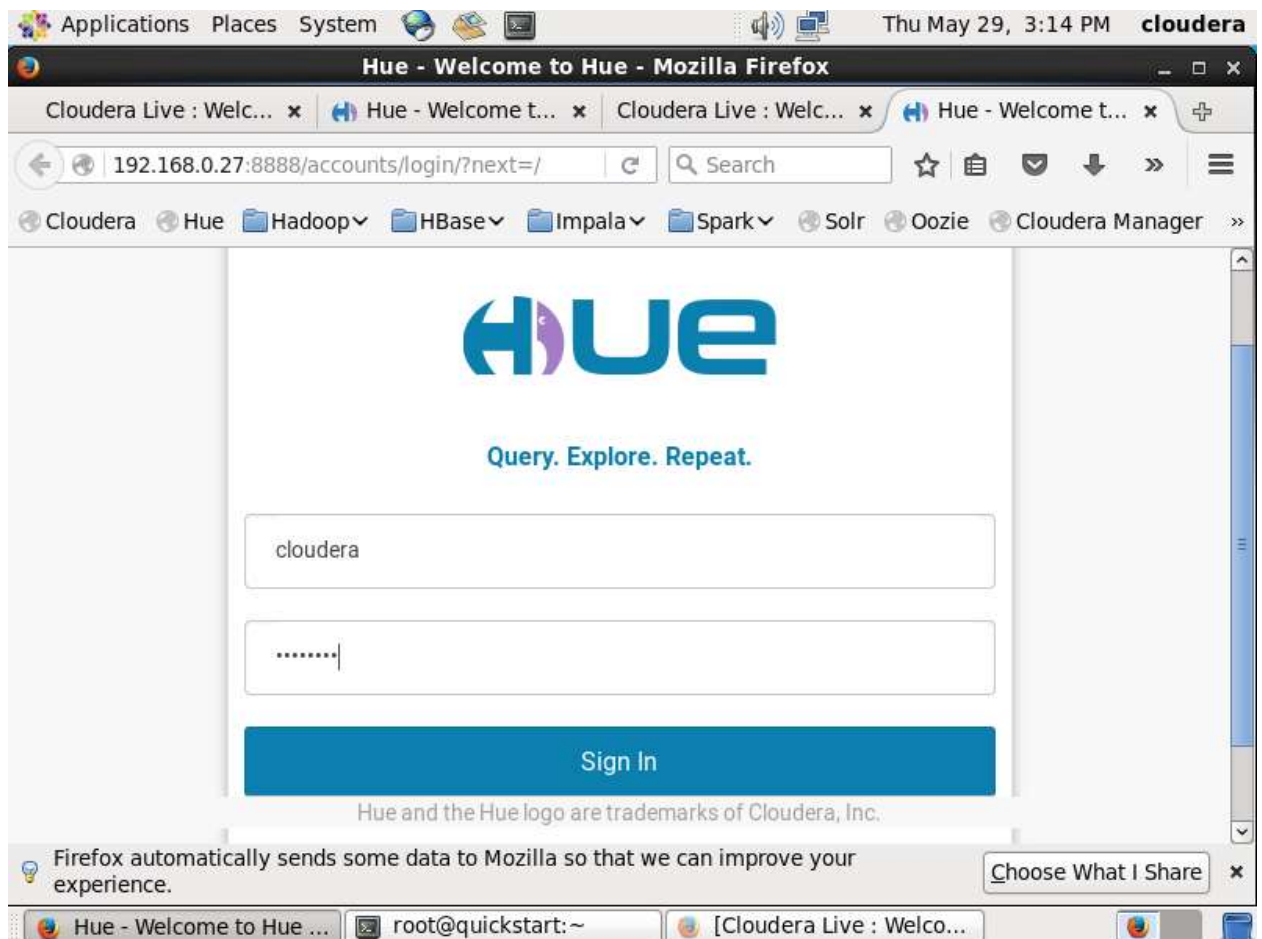
Manage Your Cluster

Cloudera Manager provides end-to-end system management for

Firefox automatically sends some data to Mozilla so that we can improve your experience. root@quickstart:~ Choose What I Share

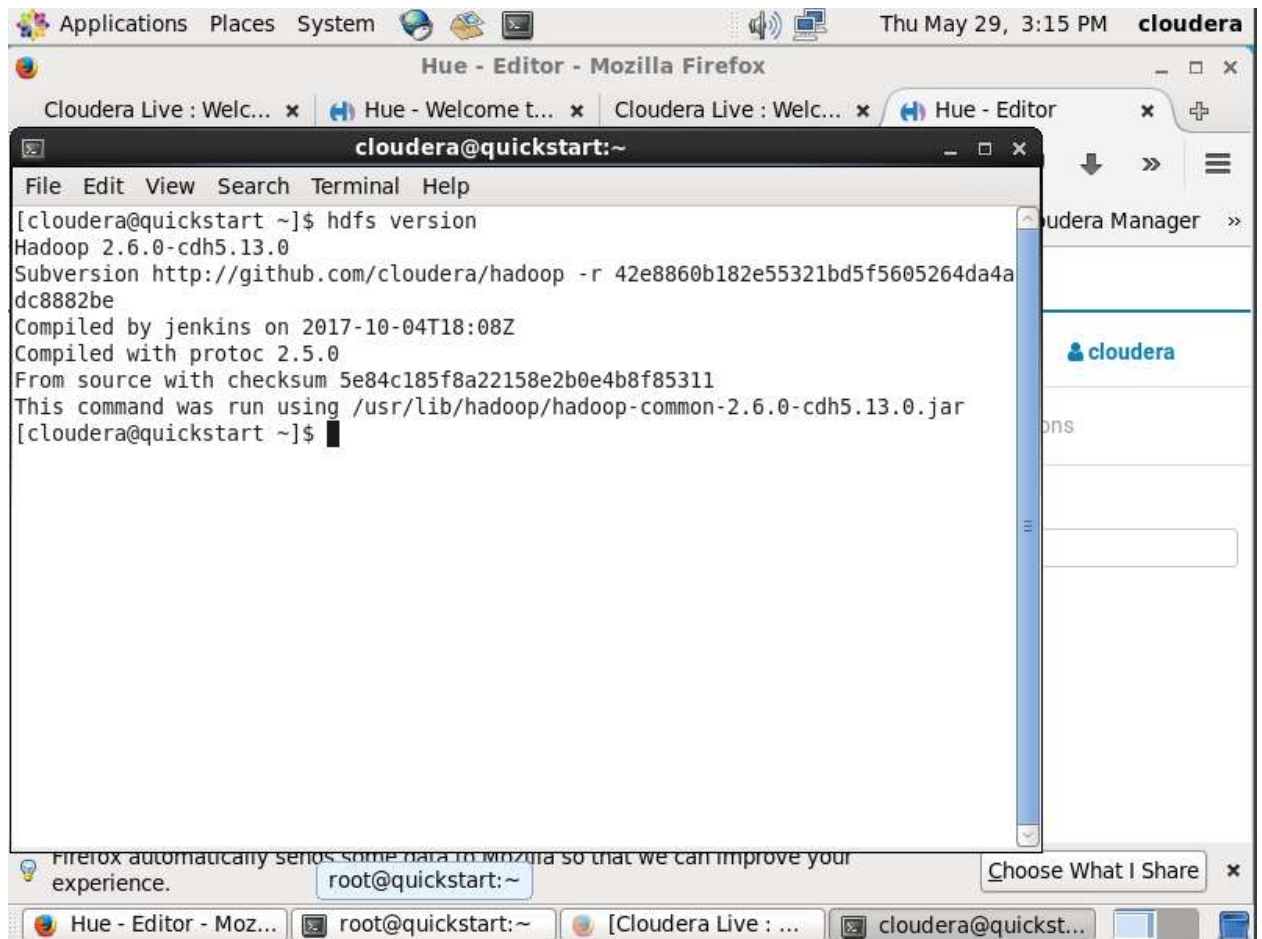
Cloudera Live : Welco... root@quickstart:~ [Cloudera Live : Welco...]

Вводим учетку cloudera cloudera



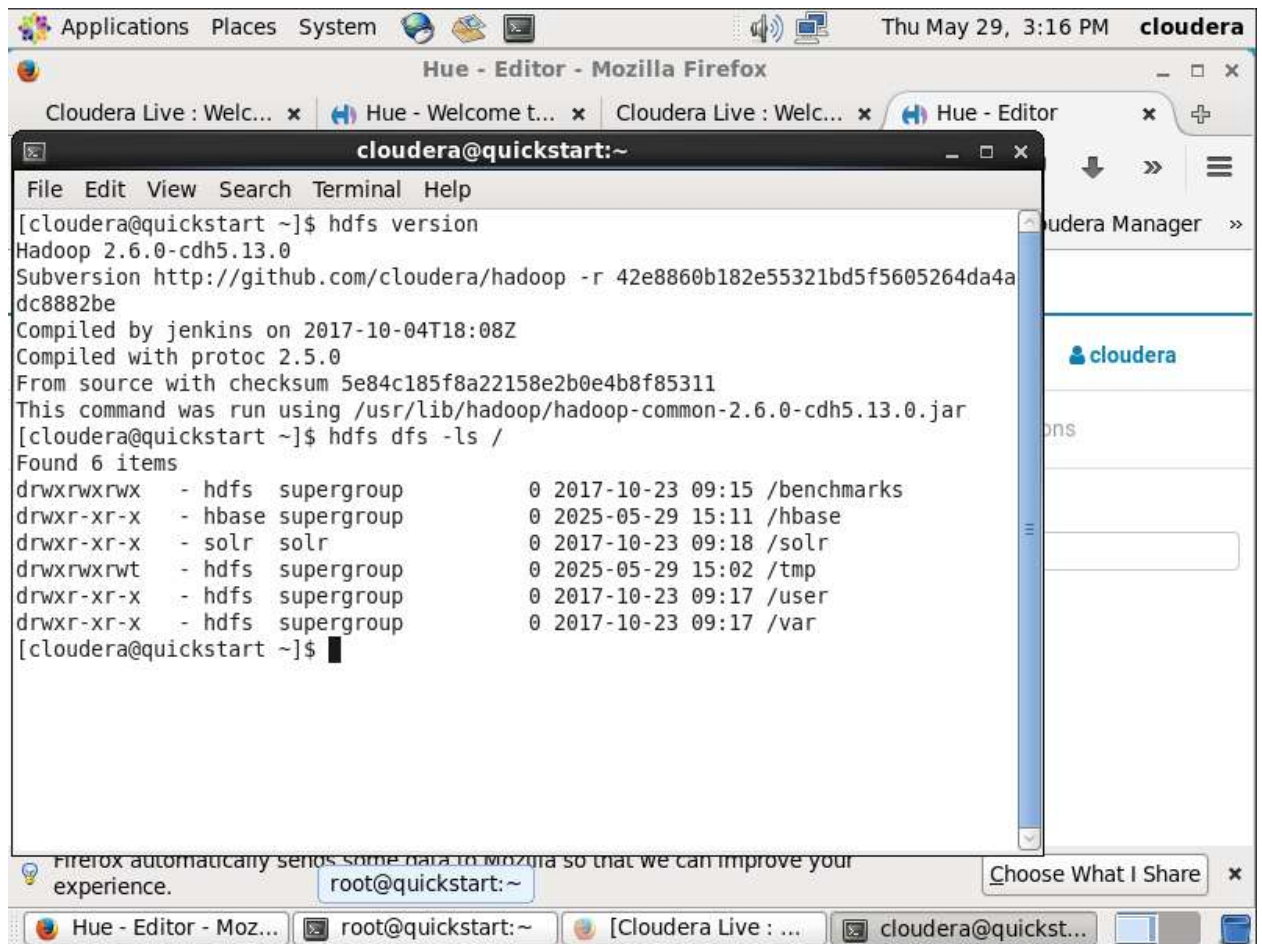
В терминале проверим версию хадуп

```
hdfs version
```



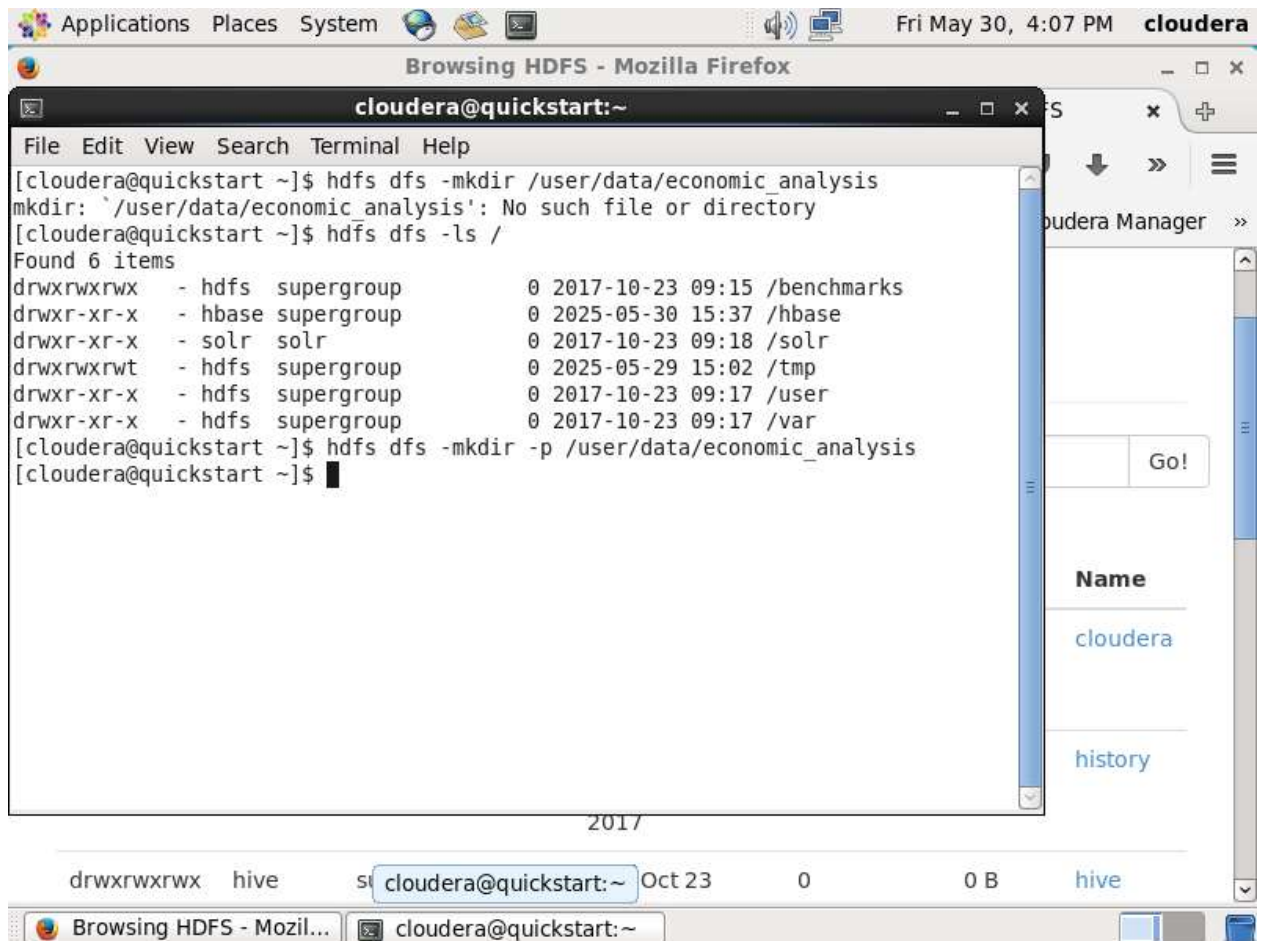
Проверим какие папки есть

```
hdfs dfs -ls /
```



Создадим папку для наших данных

```
hdfs dfs -mkdir -p /user/data/economic_analysis
```



Проверим ее в браузере в закладке Hadoop – Hdfs Namenode

Applications Places System Fri May 30, 4:07 PM cloudera

Browsing HDFS - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welc... x Hue - Editor x Browsing HDFS x

quickstart.cloudera:50070/explorer.html#/us Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager

- HDFS NameNode
- HDFS Secondary NameNode
- HDFS DataNode
- YARN ResourceManager
- YARN NodeManager

/user Open All in Tabs Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	cloudera	0 B	Mon Oct 23 09:14:53 -0700 2017	0	0 B	cloudera
drwxr-xr-x	mapred	hadoop	0 B	Mon Oct 23 09:15:52 -0700 2017	0	0 B	history
drwxrwxrwx	hive	supergroup	0 B	Mon Oct 23	0	0 B	hive

Browsing HDFS - Mozil... cloudera@quickstart:~

Перейдем в браузер файлов системы

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	cloudera	0 B	Mon Oct 23 09:14:53 -0700 2017	0	0 B	cloudera
drwxr-xr-x	mapred	hadoop	0 B	Mon Oct 23 09:15:53 -0700 2017	0	0 B	history

Скачаем файл из гитхаба

```
ssh-keygen -t ecdsa -b 256 -C vikale@mail.ru
первый вопрос - энтер
второй и третий запрос - vikale12345 (или любой другой, чтобы был сложным +-)
eval "$(ssh-agent -s)"
ssh-add ~/.ssh/id_ecdsa
вводим пароль vikale12345
теперь скопируем публичный ключ - cat ~/.ssh/id_ecdsa.pub
```


идем сюда и закидываем скопированный код <https://github.com/settings/keys>
жмем кнопку new ssh и вставляем полностью код, который нам вылез

SSH keys

New SSH key

This is a list of SSH keys associated with your account. Remove any keys that you do not recognize.

Authentication keys



hive


SHA256:C540CkTAZEM7Da0oGS/p3Lmhq2aks7Npo8IG/8rwh8Y

SSH

Added on Jun 1, 2025

Last used within the last week — Read/write

Delete



puwmy2


SHA256:8022xYSPXkX0zOTs41gNC6S10FqzZC88T5I8qna0FYV

SSH

Added on Jun 1, 2025

Last used within the last week — Read/write

Delete



hh

SHA256:BBWr/d87Kj7BXLQ5RCI+PvPWZIQ4Kf/ufgK955r9pwI

SSH

Added on Jun 1, 2025

Never used — Read/write

Delete

Вводим ssh -T [git@github.com](https://github.com)

Если все ок будет ответ —

```
remote: Counting objects: 100% (192/192), done.
remote: Compressing objects: 100% (186/186), done.
remote: Total 192 (delta 47), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (192/192), 12.66 MiB | 10.69 MiB/s, done.
Resolving deltas: 100% (47/47), done.
[cloudera@quickstart Downloads]$ ssh -T git@github.com
Hi qLuVik! You've successfully authenticated, but GitHub does not provide shell access.
[cloudera@quickstart Downloads]$ S
```

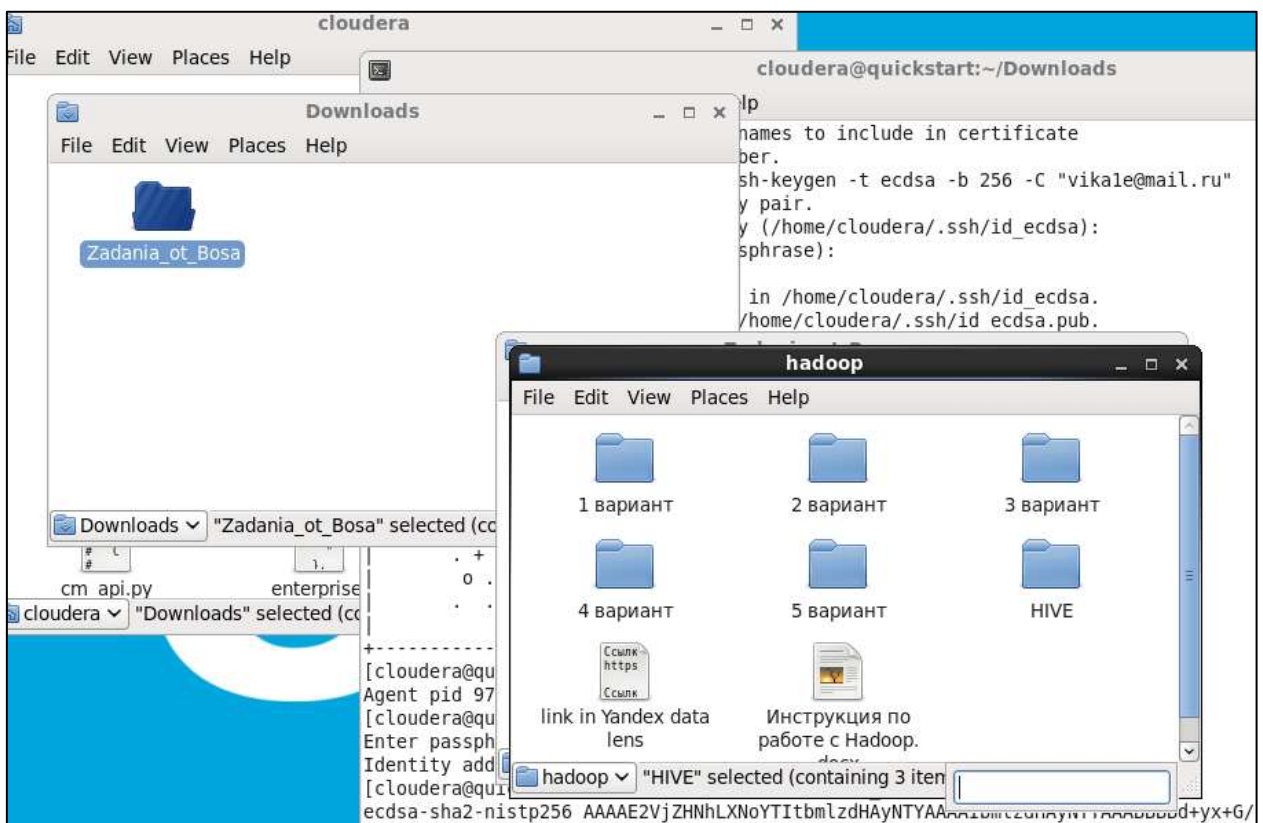
В итоге у нас загрузится директория

```

[cloudera@quickstart Downloads]$ ssh-keygen -t ecdsa -b 256 -C "vikale@mail.ru"
Generating public/private ecdsa key pair.
Enter file in which to save the key (/home/cloudera/.ssh/id_ecdsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/cloudera/.ssh/id_ecdsa.
Your public key has been saved in /home/cloudera/.ssh/id_ecdsa.pub.
The key fingerprint is:
65:bd:fb:2a:5f:1a:26:b0:16:d6:0d:55:73:71:16:08 vikale@mail.ru
The key's randomart image is:
+--[ECDSA 256]---+
|                |
|  Eo..+o*       |
|    o . +.      |
|      + .       |
|    + o .       |
|   S . o        |
|    . + .       |
|    o . + .     |
|    . o =       |
|    o+..        |
+-----+
[cloudera@quickstart Downloads]$ eval "$(ssh-agent -s)"
Agent pid 9747
[cloudera@quickstart Downloads]$ ssh-add ~/.ssh/id_ecdsa
Enter passphrase for /home/cloudera/.ssh/id_ecdsa:
Identity added: /home/cloudera/.ssh/id_ecdsa (/home/cloudera/.ssh/id_ecdsa)
[cloudera@quickstart Downloads]$ cat ~/.ssh/id_ecdsa.pub
ecdsa-sha2-nistp256 AAAAE2VjZHNhLXNoYTItbmlzdHAyNTYAAAAIbmlzdHAyNTYAAABBBBd+yx+G/b0SJkPcGagrAXdX+DCZ1tMUyjc7
XoqnIA6Fo9vCoqQ8zKptvw7LdWw/4f0qWCwt4qio8/Yc3UMPZT4= vikale@mail.ru
[cloudera@quickstart Downloads]$ git clone git@github.com:Valeria379/Zadania_ot_Bosa.git
Initialized empty Git repository in /home/cloudera/Downloads/Zadania_ot_Bosa/.git/
remote: Enumerating objects: 192, done.
remote: Counting objects: 100% (192/192), done.
remote: Compressing objects: 100% (186/186), done.
remote: Total 192 (delta 47), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (192/192), 12.66 MiB | 10.69 MiB/s, done.
Resolving deltas: 100% (47/47), done.
[cloudera@quickstart Downloads]$ █

```

Нам нужна папка HIVE я в нее закинула файлы



Теперь поехали грузить этот файл в hdfs

```
hdfs dfs -copyFromLocal
/home/cloudera/Downloads/Zadania_ot_Bosa/hadoop/HIVE/AAPL_data.csv
/user/data/economic_analysis
```

ЕСЛИ ОШИБКА

```
hdfs dfsadmin -safemode leave
```

```
[cloudera@quickstart Downloads]$ hdfs dfs -put /home/cloudera/Downloads/Zadania_ot_Bosa/hadoop/HIVE/AAPL_data.csv /user/data/economic_analysis
put: Cannot create file/user/data/economic_analysis/AAPL_data.csv._COPYING_. Name node is in safe mode.
[cloudera@quickstart Downloads]$ hdfs dfs -copyFromLocal /home/cloudera/Downloads/Zadania_ot_Bosa/hadoop/HIVE/AAPL_data.csv /user/data/economic_analysis
copyFromLocal: Cannot create file/user/data/economic_analysis/AAPL_data.csv._COPYING_. Name node is in safe mode.
[cloudera@quickstart Downloads]$ hdfs dfsadmin -safemode get
Safe mode is ON
[cloudera@quickstart Downloads]$ hdfs dfsadmin -safemode leave
Safe mode is OFF
```

Ошибка должна пройти

```
[cloudera@quickstart Downloads]$ hdfs dfs -copyFromLocal /home/cloudera/Downloads/Zadania_ot_Bosa/hadoop/HIVE/AAPL_data.csv /user/data/economic_analysis
[cloudera@quickstart Downloads]$
```

Перейдем в браузер и убедимся, что данные пришли

Browse Directory

/user/data/economic_analysis

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	supergroup	65.63 KB	Sat May 31 16:14:49 -0700 2025	1	128 MB	AAPL_data.csv

Hadoop, 2017.

Кстати файл можно загрузить руками в HUE

Cloudera Live : Welcom... * Cloudera Live : Welcom... * Hue - Editor * Cloudera Hybrid Dat... *

quickstart.cloudera:8888/hue/editor/?type=impala

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE

Query

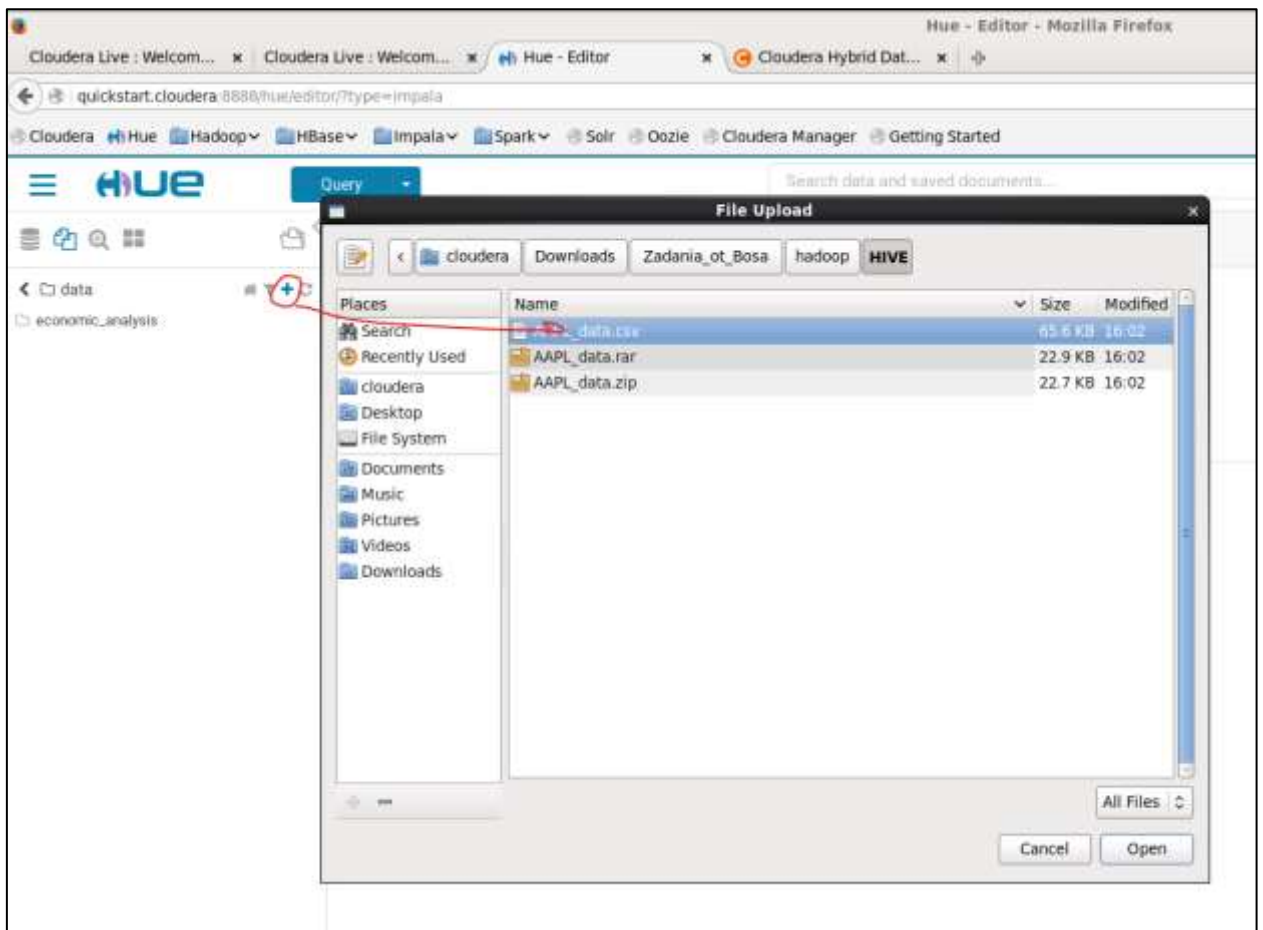
Search data and saved documents...

Impala Add a name... Add a description...

Example: SELECT * FROM tablename, or press CTRL + space

Query History Saved Queries

You don't have any saved query.



Теперь нажмем на синенькую кнопку Query и перейдем в редактор Hive, на моих скринах будет impala, но разницы нет.

Если у вас ошибка в хиве, типа такая



Тогда используйте команду

`sudo service hive-server2 restart`

```
[cloudera@quickstart ~]$ sudo service hive-server2 restart
Stopped Hive Server2: [ OK ]
Started Hive Server2 (hive-server2): [ OK ]
[cloudera@quickstart ~]$
```

Обновите страницу, либо список баз с помощью кнопки рефреш

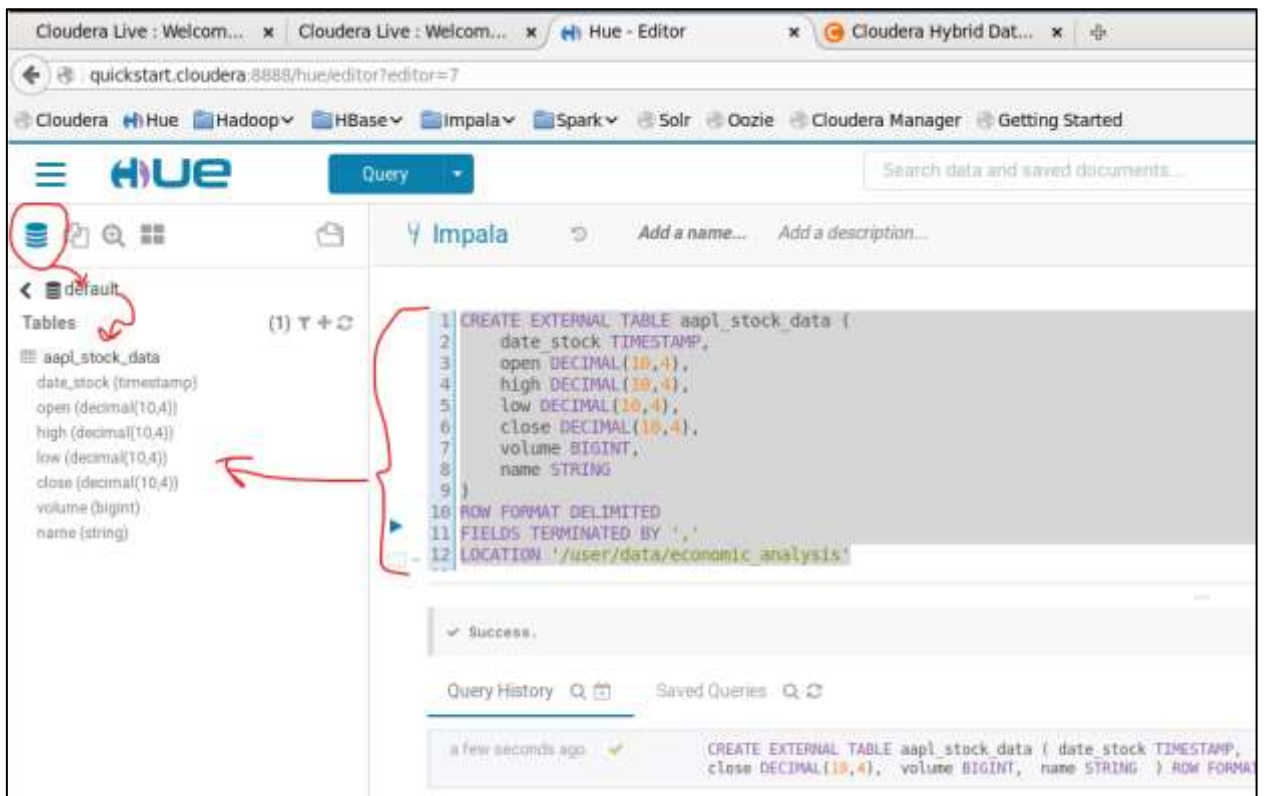
Ошибки больше не будет



Теперь создадим таблицу для этих данных, буду делать сразу тут в хью, потому что не хочу мучаться с терминалом

```
CREATE EXTERNAL TABLE aapl_stock_data (  
    date_stock DATE,  
    open DECIMAL(10,4),  
    high DECIMAL(10,4),  
    low DECIMAL(10,4),  
    close DECIMAL(10,4),  
    volume BIGINT,  
    name STRING  
)  
  
ROW FORMAT DELIMITED  
  
FIELDS TERMINATED BY ','  
  
LOCATION '/user/data/economic_analysis'  
  
TBLPROPERTIES ("skip.header.line.count"="1");
```

Таблица создана, чтобы проверить переключимся на бочонок (базы данных) и выберем нашу базу данных, раскроем столбцы



П.С. На этом скрине нет последней строки, которую юзает Босенко. Она нужна для пропуска названий столбцов в сиезви

6. Создание Внешней таблицы **title_basics** для файла **title.basics.tsv** в Hive:

```
Hive > CREATE EXTERNAL TABLE IF NOT EXISTS title_basics (
  country STRING,
  title_type STRING,
  primary_title STRING,
  original_title STRING,
  is_adult DECIMAL(1,0),
  start_year DECIMAL(4,0),
  end_year STRING,
  runtime_minutes INT,
  genres STRING
) COMMENT 'IMDb Movies' ROW FORMAT DELIMITED FIELDS TERMINATED BY
'\t' STORED AS TEXTFILE LOCATION '/user/hadoop/1900/11/16/19001116'
TBLPROPERTIES ('skip.header.line.count'='1');
```

Выведем содержимое

select * FROM aapl_stock_data limit 3


```

1 drop table if exists aapl_stock_data
2
3 CREATE EXTERNAL TABLE aapl_stock_data (
4     date_stock DATE,
5     open DECIMAL(10,4),
6     high DECIMAL(10,4),
7     low DECIMAL(10,4),
8     close DECIMAL(10,4),
9     volume BIGINT,
10    name STRING
11 )
12 ROW FORMAT DELIMITED
13 FIELDS TERMINATED BY ','
14 LOCATION '/user/data/economic_analysis'
15 TBLPROPERTIES ("skip.header.line.count"="1");
16
17 select * from aapl_stock_data limit 3

```

	aapl_stock_data.date_stock	aapl_stock_data.open	aapl_stock_data.high	aapl_stock_data.low	aapl_stock_data.close	aapl_stock_data.volume	aapl_stock_data.name
1	2013-02-08	67.7142	68.4014	66.8928	67.8542	138168416	AAPL
2	2013-02-11	68.0714	69.2771	67.6071	68.5614	129029425	AAPL
3	2013-02-12	68.5014	68.9114	66.8205	66.8428	151829363	AAPL

Все гуд, осталось только сделать запросы для аналитики

1. Все статистические показатели (экстремальные и средние)

-- Средние и экстремальные значения цен за весь период

SELECT

MIN(open) AS min_open_price,
MAX(open) AS max_open_price,
AVG(open) AS avg_open_price,
MIN(close) AS min_close_price,
MAX(close) AS max_close_price,
AVG(close) AS avg_close_price,
MIN(volume) AS min_volume,
MAX(volume) AS max_volume,
AVG(volume) AS avg_volume

FROM aapl_stock_data;

```

12 ROW FORMAT DELIMITED
13 FIELDS TERMINATED BY ','
14 LOCATION '/user/data/economic_analysis'
15 TBLPROPERTIES ("skip.header.line.count"="1");
16
17 select * from aapl_stock_data limit 3

```

```

18
19
20
21 SELECT
22     MIN(open) AS min_open_price,
23     MAX(open) AS max_open_price,
24     AVG(open) AS avg_open_price,
25     MIN(close) AS min_close_price,
26     MAX(close) AS max_close_price,
27     AVG(close) AS avg_close_price,
28     MIN(volume) AS min_volume,
29     MAX(volume) AS max_volume,
30     AVG(volume) AS avg_volume
31 FROM aapl_stock_data;

```

	min_open_price	max_open_price	avg_open_price	min_close_price	max_close_price	avg_close_price	min_volume	max_volume	avg_volume
1	66.8205	69.2771	109.05042081	66.8096	179.26	109.0666849	11479022	20483081	58327891.75504396

2. Анализ по годам

SELECT

YEAR(TO_DATE(date_stock)) AS year,

MIN(close) AS min_price,

MAX(close) AS max_price,

AVG(close) AS avg_price,

SUM(volume) AS total_volume

FROM aapl_stock_data

GROUP BY YEAR(TO_DATE(date_stock))

ORDER BY year;

```
32 -- Подбор статистики (выход значений год за год)
33 SELECT
34     YEAR(TO_DATE(date_stock)) AS year,
35     MIN(close) AS min_price,
36     MAX(close) AS max_price,
37     ROUND(AVG(close), 2) AS avg_price,
38     SUM(volume) AS total_volume
39 FROM aapl_stock_data
40 GROUP BY YEAR(TO_DATE(date_stock))
41 ORDER BY year;
```

Query History

Saved Queries

Results (6)

COLUMNS (5)

year

INT_TYPE

min_price

DECIMAL_TYPE

max_price

DECIMAL_TYPE

avg_price

DECIMAL_TYPE

total_volume

BIGINT_TYPE

	year	min_price	max_price	avg_price	total_volume
1	2013	55.7899	81.4413	67.24	21574452606
2	2014	71.3974	119	92.26	15934013092
3	2015	103.12	133	120.04	19066421452
4	2016	90.34	118.25	104.6	9686251480
5	2017	116.02	176.42	150.59	6797072145
6	2018	156.49	179.26	171.59	986094992

Все запросы будут в отдельном файле

Чтобы визуализировать:

1. Выполнить запрос
2. Нажать на кнопку графика в левом меню от вывода ответа на запрос
3. Выбрать необходимые для визуализации данные

Например, выведем на график минимальные и максимальные цены закрытия по годам

```

32 -- Таблица статистики (средняя цена, год и др.)
33 SELECT
34     YEAR(TO_DATE(date_stock)) AS year,
35     MIN(close) AS min_price,
36     MAX(close) AS max_price,
37     ROUND(AVG(close), 2) AS avg_price,
38     SUM(volume) AS total_volume
39 FROM appl_stock_data
40 GROUP BY YEAR(TO_DATE(date_stock))
41 ORDER BY year;

```

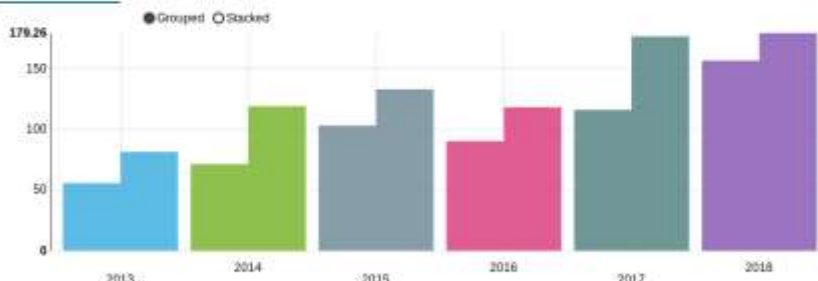
Query History Saved Queries Results (6)

☒ year
☐ min_price
☒ max_price
☐ avg_price
☐ total_volume

GROUP: Choose a column to pivot...

LIMIT: Limit the number of results to...

SORTING:



harbour