

HDFS

- 1) Выбрали датасет, загрузили на локальную машину, перенесли в hdfs и посмотрели структуру файла

```
docker-hadoop — com.docker.cli • docker exec -it namenode bash — 212x61
root@f37c326c7aad:/# hdfs dfs -mkdir -p input
root@f37c326c7aad:/# hdfs dfs -put ./home/cacao.csv input
2021-10-30 11:29:19,894 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@f37c326c7aad:/# hdfs dfs -ls
Found 1 items
drwxr-xr-x - root supergroup 0 2021-10-30 11:29 input/
root@f37c326c7aad:/# hdfs dfs -ls input/
Found 1 items
-rw-r--r-- 3 root supergroup 130866 2021-10-30 11:29 input/cacao.csv
root@f37c326c7aad:/# hdfs dfs -cat input/cacao.csv | head -5
2021-10-30 11:39:17,870 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
A. Morin;Agua Grande;1876;2016;63%;France;3.75;-t;Sao Tome
A. Morin;Kprime;1676;2015;70%;France;2.75;-t;Togo
A. Morin;Atsane;1676;2015;70%;France;3;-t;Togo
A. Morin;Akata;1680;2015;70%;France;3.5;-t;Togo
A. Morin;Quilla;1704;2015;70%;France;3.5;-t;Peru
cat: Unable to write to output stream.
```

- 2) Разделили на несколько файлов и также перенесли в hdfs

```
root@f37c326c7aad:/home# split -l 700 cacao.csv
root@f37c326c7aad:/home# ls
cacao.csv xaa xab xac
root@f37c326c7aad:/home# hdfs dfs -put * input
put: 'input/cacao.csv': File exists
2021-10-30 12:14:03,798 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-30 12:14:03,951 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-10-30 12:14:03,982 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
```

```
root@f37c326c7aad:/# hdfs dfs -ls input/
Found 4 items
-rw-r--r-- 3 root supergroup 130866 2021-10-30 11:29 input/cacao.csv
-rw-r--r-- 3 root supergroup 50159 2021-10-30 12:14 input/xaa
-rw-r--r-- 3 root supergroup 51925 2021-10-30 12:14 input/xab
-rw-r--r-- 3 root supergroup 28782 2021-10-30 12:14 input/xac
root@f37c326c7aad:/#
```

- 3) Поиграться с web-интерфейсом
Посмотрели, что файлики лежат в нужной директории hdfs

Browse Directory

Show entries Search:

| <input type="checkbox"/> | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name | <input type="checkbox"/> |
|--------------------------|------------|-------|------------|----------|---------------|-------------|------------|-----------|--------------------------|
| <input type="checkbox"/> | -rw-r--r-- | root | supergroup | 127.8 KB | Oct 30 14:29 | 3 | 128 MB | cacao.csv | |
| <input type="checkbox"/> | -rw-r--r-- | root | supergroup | 48.98 KB | Oct 30 15:14 | 3 | 128 MB | xaa | |
| <input type="checkbox"/> | -rw-r--r-- | root | supergroup | 50.71 KB | Oct 30 15:14 | 3 | 128 MB | xab | |
| <input type="checkbox"/> | -rw-r--r-- | root | supergroup | 28.11 KB | Oct 30 15:14 | 3 | 128 MB | xac | |

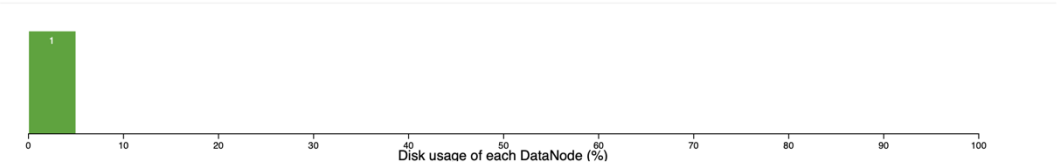
Showing 1 to 4 of 4 entries Previous **1** Next

Hadoop, 2019.

Посмотрели общую информацию о Datanode

🟢 Entering Maintenance 🟡 In Maintenance 🔴 In Maintenance & dead

Datanode usage histogram



In operation

Show 25 entries

Search:

| Node | Http Address | Last contact | Last Block Report | Capacity | Blocks | Block pool used | Version |
|--|---|--------------|-------------------|---------------------------------|--------|-----------------|---------|
| ✓ 5917a32f104a:9866 (172.18.0.3:9866) | http://5917a32f104a:9866 | 1s | 78m | 58.42 GB <div><div></div></div> | 9 | 356 KB (0%) | 3.2.1 |

Showing 1 to 1 of 1 entries

Previous 1 Next