

Скачиваем файлы образа и собираем его, выполнив команду, находясь внутри директории:

`docker build -t hive-hadoop .`

ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР

Смотрим статус NameNode: <http://localhost:50070>

Смотрим статус Datanode Information: <http://localhost:50070>

Namenode information — Mozilla Firefox

Namenode information x Marks Tools Help

Namenode information x Namenode information + localhost:50070/dfshealth.html#tab-da

## Hadoop

- Overview
- Datanodes**
- Datanode Volume Failures
- Snapshot
- Startup Progress
- Utilities -

# Datanode Information

✓ In service 
 ❗ Down 
 ⚡ Decommissioned 
 🔄 Decommissioned & dead 
 🔧 In Maintenance & dead

## Datanode usage histogram

The histogram shows a single bar at 0% disk usage, indicating that all DataNodes are currently at 0% capacity.

Disk Usage (%)	Count
0	1

## In operation

Show  entries Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓ gbhive:50010 (127.0.0.1:50010)	<a href="http://gbhive:50075">http://gbhive:50075</a>	2s	4m	514.9 GB <div><div style="width: 0%;"></div></div>	0	24 KB (0%)	2.10.1

Showing 1 to 1 of 1 entries

Previous **1** Next

## Entering Maintenance

Смотрим статус Startup Progress: <http://localhost:50070>

Namenode information — Mozilla Firefox

Namenode information × Namenode information × Namenode information × +

localhost:50070/dfshealth.html#tab-startup-progress

## Hadoop

Overview Datanodes Datanode Volume Failures Snapshot **Startup Progress**

Utilities ▾

# Startup Progress

Elapsed Time: 0 sec, Percent Complete: 100%

Phase	Completion	Elapsed Time
<b>Loading fsimage /tmp/hadoop-hduser/dfs/name/current/fsimage_00000000000000000000 325 B</b>	<b>100%</b>	
delegation tokens (0/0)	100%	
cache pools (0/0)	100%	
<b>Loading edits</b>	<b>100%</b>	<b>0 sec</b>
<b>Saving checkpoint</b>	<b>100%</b>	<b>0 sec</b>
<b>Safe mode</b>	<b>100%</b>	<b>0 sec</b>
awaiting reported blocks (0/0)	100%	

Hadoop,  
2020.

ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР  
Смотрим Browse Directory: <http://localhost:50070>

Browsing HDFS — Mozilla Firefox




Namenode info: Namenode info: arks Tools Help

Namenode info: Namenode info: localhost:50070/explorer.html#/




90%

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/ Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxrwxrwx	hduser	supergroup	0 B	Sep 13 22:30	0	0 B	logs	
<input type="checkbox"/>	drwxrwxrwx	hduser	supergroup	0 B	Sep 13 22:44	0	0 B	tmp	
<input type="checkbox"/>	drwxr-xr-x	hduser	supergroup	0 B	Sep 13 22:30	0	0 B	user	

Showing 1 to 3 of 3 entries

Previous 1 Next

Hadoop, 2020.

ПОДНИМАЕМ ЛОКАЛЬНЫЙ КЛАСТЕР

Смотрим статус DataNode: <http://localhost:50075>

DataNode Information — Mozilla Firefox

Namenode information × Namenode information × Namenode information × Namenode information × DataNode Information × +

localhost:50075/datanode.html

Hadoop Overview Utilities

## DataNode on gbhive:50010

Cluster ID:	CID-d96eb90d-cafb-423c-952c-5fc7aa5479e2
Version:	2.10.1

## Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-1035435261-172.17.0.2-1631561398162	RUNNING	0s	9 minutes	0 B (64 MB)

## Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/tmp/hadoop-hduser/dfs/data/current	DISK	24 KB	275.53 GB	0 B	0 B	0


Hadoop, 2020.

Смотрим статус NodeManager: <http://localhost:8042>

Mozilla Firefox

Namenode informa... X Namenode informa... X Namenode informa... X Namenode informa... X DataNode Informa... X localhost:8042/ X +

localhost:8042/node



ResourceManager

NodeManager

Node Information

List of Applications

List of Applications

List of Containers

Tools

NodeManager information

NodeManager information

Total Vmem allocated for Containers 8.40 GB

Vmem enforcement enabled true

Vmem enforcement enabled true

Total Pmem allocated for Container 4 GB

Pmem enforcement enabled true

Total Vcores allocated for Containers 8

NodeHealthyStatus true

LastNodeHealthTime Mon Sep 13 19:40:16 GMT 2021

NodeHealthReport

NodeManager started on Mon Sep 13 19:30:15 GMT 2021

NodeManager Version: 2.10.1 from 1827467c9a56f133025f28557bfc2c562d78e816 by centos source checksum 2da9946ffe56799794b77621fbe0be1a on 2020-09-14T13:24Z

Hadoop Version: 2.10.1 from 1827467c9a56f133025f28557bfc2c562d78e816 by centos source checksum 3114edef868f1f3824e7d0f68be03650 on 2020-09-14T13:17Z

Смотрим статус ResourceManager: <http://localhost:8088>


All Applications — Mozilla Firefox

Namenode information X Namenode information X Namenode information X Namenode information X DataNode Information X localhost:8042/ X All Applications X +

localhost:8088/cluster

90%

LoggedIn in as: dr who



All Applications

Cluster

About

Nodes

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Cluster Metrics				Cluster Metrics				Cluster Metrics				Cluster Metrics			
Apps Submitted		Apps Pending		Apps Running		Apps Completed		Containers Running		Used Resources		Total Resources			
0	0	0	0	0	0	0	0	0	0	0	0	0			

Cluster Nodes Metrics

Active Nodes		Decommissioning Nodes		Decommissioned Nodes		Lost Nodes		Unhealthy Nodes		Rebooted Nodes		Shutdown Nodes	
1	0	0	0	0	0	0	0	0	0	0	0	0	

Scheduler Metrics

Scheduler Metrics				Scheduler Metrics				Scheduler Metrics			
Capacity Scheduler				Scheduling Resource Type				Minimum Allocation			
[<name=memory-mb default-unit=Mi type=COUNTABLE>, <name=vcores default-unit= type=COUNTABLE>]				<memory:256, vCores:1>				<memory:1024, vCores:4>			

Showing 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Allocated GPUs	Reserved CPU Vcores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																						

Showing 0 to 0 of 0 entries

First Previous Next Last

ПРОВЕРЯЕМ РАБОТУ HIVE

Выведем список таблиц в hive:

```
Терминал - hduser@gbhive: ~
Файл  Правка  Вид  Терминал  Вкладки  Справка
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts
.
0.0.0.0: starting secondarynamenode, logging to /home/hduser/hadoop/logs/hadoop-
hduser-secondarynamenode-gbhive.out
starting yarn daemons
starting resource manager, logging to /home/hduser/hadoop/logs/yarn--resourcemana
ger-gbhive.out
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known h
osts.
localhost: starting nodemanager, logging to /home/hduser/hadoop/logs/yarn-hduser
-nodemanager-gbhive.out
hduser@gbhive:~$ hive -e 'show tables;'
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hduser/hive/lib/log4j-slf4j-impl-2.6.2.j
ar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hduser/hadoop/share/hadoop/common/lib/sl
f4j-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]

Logging initialized using configuration in jar:file:/home/hduser/hive/lib/hive-c
ommon-2.3.9.jar!/hive-log4j2.properties Async: true
OK
Time taken: 20.176 seconds
```

## РАБОТА С CRON

Добавляем в систему:

```
sudo apt-get install cron
```

Включаем:

```
sudo service cron start
```

Открываем для редактирования:

```
crontab -e
```

Добавляем правила:

```
* * * * * echo `date` >> /home/hduser/datus.log
```

```
0 * * * * hdfs dfs -put /home/hduser/datus.log "/user/hduser/"`date +%s`.log"
```

```
0 * * * * wget -q mail.ru -O /home/hduser/mail-ru.html
```

Смотрим существующие правила:

```
crontab -l
```



```
Терминал - igor@igor-MS-7808: ~/Настройка потоков данных. Apache Airflo
Файл  Правка  Вид  Терминал  Вкладки  Справка

проблемы зависимостей – оставляем не настроенным
При обработке следующих пакетов произошли ошибки:
python3-lib2to3
python3-distutils
libglib2.0-dev-bin
libglib2.0-dev:amd64
libharfbuzz-dev:amd64
libicu-le-hb-dev:amd64
libicu-dev
libxml2-dev:amd64
(base) igor@igor-MS-7808:~/Настройка потоков данных. Apache Airflow/1. Планирова
ние задач. Введение Apache AirFlow/hive-hdp-image$ sudo service cron start
(base) igor@igor-MS-7808:~/Настройка потоков данных. Apache Airflow/1. Планирова
ние задач. Введение Apache AirFlow/hive-hdp-image$ crontab -e
no crontab for igor - using an empty one

Select an editor. To change later, run 'select-editor'.
 1. /bin/nano          <---- easiest
 2. /usr/bin/vim.basic
 3. /usr/bin/vim.tiny
 4. /usr/bin/emacs25
 5. /bin/ed

Choose 1-5 [1]: █
```

Запуск команды каждые 4 часа:

```
0 */4 * * * ./run.sh
```

Задачи определяют «что запускать?», а операторы — «как запустить?»