

Команды для установки Hadoop.

Обновление системы.

1. `sudo apt update`
2. `sudo apt list --upgradable`

Установка и проверка OpenJDK 8.

3. `sudo apt install openjdk-8-jdk -y`
4. `java -version`
5. `javac -version`

Установка сервера и клиента OpenSSH.

6. `sudo apt install openssh-server`
7. `sudo apt install openssh-client` (клиент установлен по умолчанию)

Создание нового пользователя Hadoop и настройка доступа.

8. `sudo adduser hdfs` (Просит заполнить пароль, имя пользователя и дополнительную информацию)
9. `su - hdfs`
10. `ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa`
11. `cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys`
12. `chmod 0600 ~/.ssh/authorized_keys`

Скачивание и распаковка hadoop

13. `wget https://apache-mirror.rbc.ru/pub/apache/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz`
14. `tar xzf hadoop-3.2.1.tar.gz`

Редактирование .bashrc

15. `emacs .bashrc`

Текст для вставки. Вставляем в конец файла и сохраняем.

`#Hadoop Related Options`

```
export HADOOP_HOME=/home/hdfs/hadoop-3.2.1
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

16. `source ~/.bashrc`

Редактирование hadoop-env.sh

17. `emacs $HADOOP_HOME/etc/hadoop/hadoop-env.sh`

Текст для вставки

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

Редактирование core-site.xml

18. `emacs $HADOOP_HOME/etc/hadoop/core-site.xml`

Текст для вставки

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/home/hadoop/tmpdata</value>
```

```
</property>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://127.0.0.1:9000</value>
  </property>
</configuration>
```

Создаем tmp катлог указанный в файле core-site.xml.

19. `mkdir tmpdata`

Редактирование hdfs-site.xml

20. `emacs $HADOOP_HOME/etc/hadoop/hdfs-site.xml`

Текст для вставки

```
<configuration>
  <property>
    <name>dfs.data.dir</name>
    <value>/home/hdfs/dfsdata/namenode</value>
  </property>
  <property>
    <name>dfs.data.dir</name>
    <value>/home/hdfs/dfsdata/datanode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

Создаем катлоги указанные в файле hdfs-site.xml.

21. `mkdir dfsdata`

22. `mkdir dfsdata/namenode dfsdata/datanode`

Редактирование mapred-site.xml

23. `emacs $HADOOP_HOME/etc/hadoop/mapred-site.xml`

Текст для вставки

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

Редактирование yarn-site.xml

23. `emacs $HADOOP_HOME/etc/hadoop/yarn-site.xml`

Текст для вставки

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

```

</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HA-
DOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PER-
PEND_DISTCACHE,HADOOP_YARN_HOME,HA-
DOOP_MAPRED_HOME</value>
</property>
</configuration>

```

Форматирование HDFS NameNode

24. `hdfs namenode -format`

Запуск hadoop кластера

25. `~/hadoop-3.2.1/sbin/start-all.sh`

```

hdfs@iavode-ubuntu:~$ ~/hadoop-3.2.1/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdfs in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [iavode-ubuntu]
Starting resourcemanager
Starting nodemanagers

```

Проверка активности всех демонов

26. `jps`

```

hdfs@iavode-ubuntu:~$ jps
6947 ResourceManager
6563 DataNode
6762 SecondaryNameNode
7274 NodeManager
6395 NameNode
7435 Jps

```

Работа на hadoop кластере

Создаем тестовую папку

27. `hdfs dfs -mkdir /odegov`

Проверяем ее наличие

28. `hdfs dfs -ls /`

```

hdfs@iavode-ubuntu:~$ hdfs dfs -mkdir /odegov
hdfs@iavode-ubuntu:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x  - hdfs supergroup          0 2020-10-16 11:19 /odegov
hdfs@iavode-ubuntu:~$

```

Заняск WebUI

29. В адресной строке браузера вводим: *http://localhost:9870*

Activities

Firefox Web Browser

OKT 16 11:22

en

🔊 🔌 🔍

Namenode Information - Mozilla Firefox

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (active)

Started:	Fri Oct 16 11:15:48 +0300 2020
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 18:56:00 +0300 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-d0a5d66d-2ed3-41b7-b51a-a5fb42a2d849
Block Pool ID:	BP-1260690512-127.0.1.1-1602834748775

Summary

Security is off.
Safemode is off.
2 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 2 total filesystem object(s).
Heap Memory used 246.48 MB of 442.5 MB Heap Memory. Max Heap Memory is 3.46 GB.
Non Heap Memory used 47.33 MB of 48.75 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	127.6 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	24.4 GB
DFS Remaining:	96.65 GB (75.74%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)

Activities

Firefox Web Browser

OKT 16 11:25

ru

🔊 🔌 🔍

Namenode Information - Mozilla Firefox

localhost:9870/dfshealth.html#tab-overview

Summary

Security is off.
Safemode is off.
2 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 2 total filesystem object(s).
Heap Memory used 258.56 MB of 442.5 MB Heap Memory. Max Heap Memory is 3.46 GB.
Non Heap Memory used 48.72 MB of 50 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	127.6 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	24.41 GB
DFS Remaining:	96.65 GB (75.74%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Fri Oct 16 11:15:48 +0300 2020
Last Checkpoint Time	Fri Oct 16 10:52:28 +0300 2020
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Activities Firefox Web Browser

OKT 16 11:25

ru

namenode information

localhost:9870/dfshealth.html#tab-overview

Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Fri Oct 16 11:15:48 +0300 2020
Last Checkpoint Time	Fri Oct 16 10:52:28 +0300 2020
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 5

Journal Manager	State
FileJournalManager(root=/home/hdfs/tmpdata/dfs/name)	EditLogFileOutputStream(/home/hdfs/tmpdata/dfs/name/current/edits_inprogress_00000000000000000002)

NameNode Storage

Storage Directory	Type	State
/home/hdfs/tmpdata/dfs/name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	127.6 GB	24 KB (0%)	96.65 GB (75.74%)	24 KB	1

Hadoop, 2019.

Activities Firefox Web Browser

OKT 16 11:24

ru

namenode information

localhost:9870/dfshealth.html#tab-datanode

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Datanode Information

✓ In service ⚠ Down ⚡ Decommissioning ⚡ Decommissioned ⚡ Decommissioned & dead
➡ Entering Maintenance ➡ In Maintenance ➡ In Maintenance & dead

Datanode usage histogram

Disk usage of each DataNode (%)

In operation

Show 25 entries Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓ lavode-ubuntu:9866 (127.0.0.1:9866)	http://lavode-ubuntu:9864	0s	8m	127.6 GB	0	24 KB (0%)	3.2.1

Showing 1 to 1 of 1 entries Previous 1 Next

Entering Maintenance

Доступ к DataNode: <http://localhost:9864>

DataNode on iavode-ubuntu:9866

Cluster ID:	CID-d0a5d66d-2ed3-41b7-b51a-a5fb42a2d849
Version:	3.2.1, rb3cbb467e22ea829b3808f4b7b01d07e0bf3842

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-1260690512-127.0.1.1-1602834748775	RUNNING	0s	10 minutes	0 B (64 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/hdfs/dfsdata/datanode	DISK	32 KB	96.65 GB	0 B	0 B	0

Hadoop, 2019.

Доступ к YARN: <http://localhost:8088>

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=Mb), vcores]	<memory:1024, vcores:1>	<memory:8192, vcores:4>	0

Show 20 entries

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

Showing 0 to 0 of 0 entries

Завершение работы

29. `~/hadoop-3.2.1/sbin/stop-all.sh`

```
hdfs@iavode-ubuntu:~$ ~/hadoop-3.2.1/sbin/stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hdfs in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [iavode-ubuntu]
Stopping nodemanagers
Stopping resourcemanager
```