

Установка Java

`sudo apt update`

`$ java -version`

`~$ sudo apt install openjdk-13-jre-headless`

`~$ sudo apt install openjdk-13-jre-headless`

проверка:

`~$ java -version`

Добавляем группу и пользователя для управления Hadoop

`$ sudo addgroup hadoop`

`~$ sudo adduser --ingroup hadoop hduser`

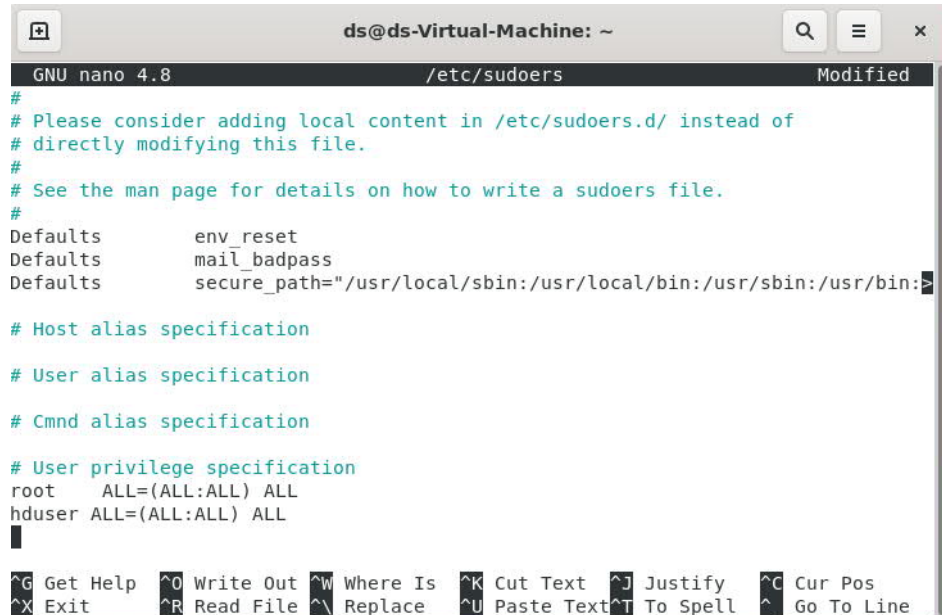
Устанавливаем пароль на пользователя pwd: 123456

Настроиваем права доступа для пользователя `hduser`

`$ sudo apt-get install nano`

`~$ sudo nano /etc/sudoers`

Вписываем строку `hduser ALL=(ALL:ALL) ALL`



```
ds@ds-Virtual-Machine: ~
GNU nano 4.8 /etc/sudoers Modified
#
# Please consider adding local content in /etc/sudoers.d/ instead of
# directly modifying this file.
#
# See the man page for details on how to write a sudoers file.
#
Defaults        env_reset
Defaults        mail_badpass
Defaults        secure_path="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:
# Host alias specification

# User alias specification

# Cmnd alias specification

# User privilege specification
root    ALL=(ALL:ALL) ALL
hduser  ALL=(ALL:ALL) ALL
^G Get Help  ^O Write Out ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File ^\ Replace   ^U Paste Text ^T To Spell  ^_ Go To Line
```

Сохраняем и закрываем документ CTRL+O CTRL+X

Установка SSH-сервера

`sudo apt-get install openssh-server`

V 3.0 16-06-2021

Авторизация под пользователем hduser и генерация SSH-ключей

```
$ sudo su hduser
```

```
$ ssh-keygen -t rsa -P ""
```

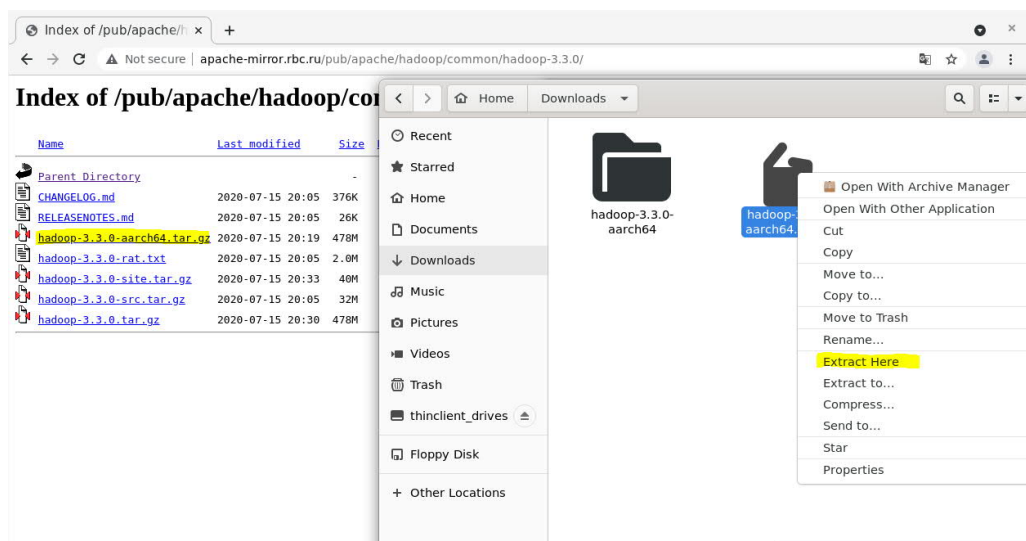
```
cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

```
ds@ds-Virtual-Machine:~$ sudo su hduser
hduser@ds-Virtual-Machine:/home/ds$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
Created directory '/home/hduser/.ssh'.
Your identification has been saved in /home/hduser/.ssh/id_rsa
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:Gv0SariELlw/bXaT1SMuSb4k5ERhdtysrCLN+d12m8c hduser@ds-Virtual-Machine
The key's randomart image is:
+---[RSA 3072]-----+
|      +..o          |
|      o o . o       |
|      .. .         |
|      . . . o       |
|    +o .o+ S .      |
|    .o.=+ . B o o   |
|    ..o.+oB.B . o   |
|    .o..*oX + ..E   |
|    .+= .. = .oo    |
+---[SHA256]-----+
hduser@ds-Virtual-Machine:/home/ds$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

Установка Hadoop

Скачиваем архив:

<http://apache-mirror.rbc.ru/pub/apache/hadoop/common/>



И распаковываем

Создание рабочей папки

Переходим в учетную запись hduser

```
~$ sudo su hduser
```

```
$ cd Downloads
```

```
$ ls
```

```
hadoop-3.3.0-aarch64 hadoop-3.3.0-aarch64.tar.gz
```

Создаем каталог сервиса Hadoop

V 3.0 16-06-2021

```
$ sudo mkdir -p /usr/local/Hadoop
```

Переносим распакованный архив в папку

```
$ sudo mv hadoop-3.3.0-aarch64 /usr/local/Hadoop/hadoop-3.3.0-aarch64
```

Создание рабочих каталогов для HDFS (NameNode и DataNode) и назначение владельца каталога hadoop (пользователь hduser):

```
$ sudo mkdir -p /usr/local/Hadoop/hadoop_tmp/hdfs/namenode
```

```
$ sudo mkdir -p /usr/local/Hadoop/hadoop_tmp/hdfs/datanode
```

Устанавливаем права на доступ

```
$ sudo chown hduser:hadoop -R /usr/local/Hadoop/
```

Настройка переменные окружения

```
cd
```

```
$ nano ~/.bashrc
```

- Переходим в конец файла и вводим текст

```
export JAVA_HOME=/usr/lib/jvm/java-1.13.0-openjdk-amd64
```

```
export HADOOP_HOME=/usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0
```

```
export PATH=$PATH:$HADOOP_HOME/bin
```

```
export PATH=$PATH:$HADOOP_HOME/sbin
```

```
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 HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

Сохраняем и закрываем документ CTRL+O CTRL+X

```
$ source ~/.bashrc
```

Настройка Hadoop

Добавить переменную JAVA_HOME в файл /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/hadoop-env.sh

```
$ nano /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/hadoop-env.sh
```

```
export JAVA_HOME=${JAVA_HOME}
```

```
export JAVA_HOME=/usr/lib/jvm/java-1.13.0-openjdk-amd64
```

```
GNU nano 4.8 /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/hadoop-env.sh

###
# Registry DNS specific parameters
###
# For privileged registry DNS, user to run as after dropping privileges
# This will replace the hadoop.id.str Java property in secure mode.
# export HADOOP_REGISTRYDNS_SECURE_USER=yarn

# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="-jvm server"
export JAVA_HOME=${JAVA_HOME}

export JAVA_HOME=/usr/lib/jvm/java-1.13.0-openjdk-amd64
```

Сохраняем и закрываем документ CTRL+O CTRL+X

Настройка Hadoop 2

`sudo nano /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/core-site.xml`

Настройка файла core-site.xml. Для этого открыть файл в редакторе с помощью `<configuration>`

`<configuration>`

`<property>`

`<name>fs.default.name</name>`

`<value>hdfs://localhost:9000</value>`

`</property>`

`</configuration>`

Здесь мы указали путь к основной машине (NameNode) файловой системы HDFS, где localhost только для локальной установки HADOOP (в распределенной версии тут либо IP-адрес, либо DNS-имя); 9000 - порт, может быть иным, но лучше не менять.

```
GNU nano 4.8 /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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

Сохраняем и закрываем документ CTRL+O CTRL+X

Настройка yarn-site.xml. Для этого открыть файл в редакторе с помощью команды

`sudo nano /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/yarn-site.xml`

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>file:/usr/local/Hadoop/hadoop_tmp/hdfs/namenode</value>
</property>
<property>
<name>dfs.datanode.name.dir</name>
<value>file:/usr/local/Hadoop/hadoop_tmp/hdfs/datanode</value>
</property>
</configuration>
```

Сохраняем и закрываем документ CTRL+O CTRL+X

Настройка mapred-site.xml

`sudo nano /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/mapred-site.xml`

V 3.0 16-06-2021

```
GNU nano 4.8 /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop/mapred-site.xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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

Сохраняем и закрываем документ CTRL+O CTRL+X

Перезагружаем ОС, заходим под пользователем hduser

Запускаем сервис Hadoop, предварительно отформатировав раздел, где будут храниться данные:

`$ hdfs namenode -format`

```
hduser@ds-Virtual-Machine:~$ hdfs namenode -format
WARNING: /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/logs does not exist. Creating.
2021-06-14 15:12:35,620 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = ds-Virtual-Machine/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.0
STARTUP_MSG:   classpath = /usr/local/Hadoop/hadoop-3.3.0-aarch64/hadoop-3.3.0/etc/hadoop:/u
```

Запуск HADOOP

Запуск HADOOP заключается в запуске менеджеров HDFS и YARN. Для этого необходимо запустить два скрипта.

Для HDFS: start-dfs.sh

Для YARN: start-yarn.sh

Успешность запуска всех демонов HADOOP можно проверить с помощью команды jps.

Если все настроено правильно, должно быть запущено 5 процессов HADOOP:
NodeManager, NameNode, DataNode, ResourceManager, SecondaryNameNode.

V 3.0 16-06-2021

```
hduser@ds-Virtual-Machine:~$ start-dfs.sh
Starting namenodes on [localhost]
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [ds-Virtual-Machine]
ds-Virtual-Machine: Warning: Permanently added 'ds-virtual-machine' (ECDSA) to the list of known hosts.
2021-06-14 15:16:51,428 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
lasses where applicable
hduser@ds-Virtual-Machine:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hduser@ds-Virtual-Machine:~$ jps
3920 Jps
3441 ResourceManager
3570 NodeManager
3027 DataNode
3208 SecondaryNameNode
2891 NameNode
```