**Задача:**

Установить, настроить и запустить Hadoop Сore в минимальной конфигурации. Для этого потребуется подготовить 2 виртуальные машины: VM1 - headnode; VM2 - worker. Понимание принципов работы Hadoop и его компонентов для успешной сдачи задания не требуется.

Все инструкции и команды для каждого шага задания должны быть сохранены в файле.

**Детальная формулировка:**

1. Установить CentOS на 2 виртуальные машины:

* VM1: 2CPU, 2-4G памяти, системный диск на 15-20G и дополнительные 2 диска по 5G
* VM2: 2CPU, 2-4G памяти, системный диск на 15-20G и дополнительные 2 диска по 5G

Все дальнейшие действия будут выполняться на обеих машинах, если не сказано иначе.

```

Centos7 Type: Linux, Version: Red Hat (64bit)

BaseMemory: 2560MB, Processors: 2

HDD Type: VDI, Size: 20GB

INSTALLATION DESTINATION: ATA VBOX HARDDISK Autopartitioning

NETWORK: enp0s3, IP: 10.0.2.15

Internal Network: enp0s8, IP: 192.168.2.11

```

2. При установке CentOS создать дополнительного пользователя **exam** и настроить для него использование sudo без пароля. Все последующие действия необходимо выполнять от этого пользователя, если не указано иное.

```

USER: exam, Make admin: ON (при установке)

[root@localhost exam]# echo 'exam ALL=(ALL) NOPASSWD:ALL' >> /etc/sudoers

```

3. Установить OpenJDK8 из репозитория CentOS.

```

[exam@localhost ~]$ sudo yum -y update

[exam@localhost ~]$ sudo yum install java-1.8.0-openjdk

[exam@localhost ~]$ java -version

openjdk version "1.8.0\_312"

OpenJDK Runtime Environment (build 1.8.0\_312-b07)

OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)

```

4. Скачать архив с Hadoop версии 3.1.2 (<https://hadoop.apache.org/release/3.1.2.html>)

```

[exam@localhost ~]$ curl -O https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 317M 100 317M 0 0 247k 0 0:21:52 0:21:52 --:--:-- 384k

```

5. Распаковать содержимое архива в /opt/hadoop-3.1.2/

```

[exam@localhost ~]$ sudo tar xvzf hadoop-3.1.2.tar.gz -C /opt/

(папка hadoop-3.1.2 внутри opt создается при распаковке)

[exam@localhost ~]$ ll /opt/hadoop-3.1.2/

total 176

drwxr-xr-x. 2 1001 1002 183 Jan 28 2019 bin

drwxr-xr-x. 3 1001 1002 20 Jan 28 2019 etc

drwxr-xr-x. 2 1001 1002 106 Jan 28 2019 include

drwxr-xr-x. 3 1001 1002 20 Jan 28 2019 lib

drwxr-xr-x. 4 1001 1002 288 Jan 28 2019 libexec

-rw-rw-r--. 1 1001 1002 147145 Jan 23 2019 LICENSE.txt

-rw-rw-r--. 1 1001 1002 21867 Jan 23 2019 NOTICE.txt

-rw-rw-r--. 1 1001 1002 1366 Jan 23 2019 README.txt

drwxr-xr-x. 3 1001 1002 4096 Jan 28 2019 sbin

drwxr-xr-x. 4 1001 1002 31 Jan 28 2019 share

```

6. Сделать симлинк /usr/local/hadoop/current/ на директорию /opt/hadoop-3.1.2/

```

[exam@localhost ~]$ sudo mkdir /usr/local/hadoop/

[exam@localhost ~]$ sudo ln -s /opt/hadoop-3.1.2/ /usr/local/hadoop/current

[exam@localhost ~]$ ll /usr/local/hadoop/current/

total 176

drwxr-xr-x. 2 1001 1002 183 Jan 28 2019 bin

drwxr-xr-x. 3 1001 1002 20 Jan 28 2019 etc

drwxr-xr-x. 2 1001 1002 106 Jan 28 2019 include

drwxr-xr-x. 3 1001 1002 20 Jan 28 2019 lib

drwxr-xr-x. 4 1001 1002 288 Jan 28 2019 libexec

-rw-rw-r--. 1 1001 1002 147145 Jan 23 2019 LICENSE.txt

-rw-rw-r--. 1 1001 1002 21867 Jan 23 2019 NOTICE.txt

-rw-rw-r--. 1 1001 1002 1366 Jan 23 2019 README.txt

drwxr-xr-x. 3 1001 1002 4096 Jan 28 2019 sbin

drwxr-xr-x. 4 1001 1002 31 Jan 28 2019 share

```

7. Создать пользователей **hadoop, yarn** и **hdfs**, а также группу **hadoop**, в которую необходимо добавить всех этих пользователей

```

[exam@localhost ~]$ for user in hadoop yarn hdfs; do sudo useradd $user; done

[exam@localhost ~]$ getent passwd {1000..1005}

exam:x:1000:1000:exam:/home/exam:/bin/bash

hadoop:x:1001:1001::/home/hadoop:/bin/bash

yarn:x:1002:1002::/home/yarn:/bin/bash

hdfs:x:1003:1003::/home/hdfs:/bin/bash

[exam@localhost ~]$ sudo groupadd hadoop

groupadd: group 'hadoop' already exists

[exam@localhost ~]$ for user in hadoop yarn hdfs; do sudo usermod -aG hadoop $user; done

```

8. Создать для обоих дополнительных дисков разделы размером в 100% диска.

```

[exam@localhost ~]$ sudo gdisk /dev/sdb

GPT fdisk (gdisk) version 0.8.10

Partition table scan:

MBR: not present

BSD: not present

APM: not present

GPT: not present

Creating new GPT entries.

Command (? for help): n

Partition number (1-128, default 1):

First sector (34-10485726, default = 2048) or {+-}size{KMGTP}:

Last sector (2048-10485726, default = 10485726) or {+-}size{KMGTP}:

Current type is 'Linux filesystem'

Hex code or GUID (L to show codes, Enter = 8300): 8e00

Changed type of partition to 'Linux LVM'

Command (? for help): w

Final checks complete. About to write GPT data. THIS WILL OVERWRITE EXISTING

PARTITIONS!!

Do you want to proceed? (Y/N): Y

OK; writing new GUID partition table (GPT) to /dev/sdb.

The operation has completed successfully.

[exam@localhost ~]$ sudo gdisk /dev/sdc

GPT fdisk (gdisk) version 0.8.10

Partition table scan:

MBR: not present

BSD: not present

APM: not present

GPT: not present

Creating new GPT entries.

Command (? for help): n

Partition number (1-128, default 1):

First sector (34-10485726, default = 2048) or {+-}size{KMGTP}:

Last sector (2048-10485726, default = 10485726) or {+-}size{KMGTP}:

Current type is 'Linux filesystem'

Hex code or GUID (L to show codes, Enter = 8300): 8e00

Changed type of partition to 'Linux LVM'

Command (? for help): w

Final checks complete. About to write GPT data. THIS WILL OVERWRITE EXISTING

PARTITIONS!!

Do you want to proceed? (Y/N): Y

OK; writing new GUID partition table (GPT) to /dev/sdc.

The operation has completed successfully.

```

9. Инициализировать разделы из п.8 в качестве физических томов для LVM.

```

[exam@localhost ~]$ sudo pvcreate /dev/sdb1

[exam@localhost ~]$ sudo pvcreate /dev/sdc1

[exam@localhost ~]$ sudo pvs

PV VG Fmt Attr PSize PFree

/dev/sda2 centos lvm2 a-- <19,00g 0

/dev/sdb1 lvm2 --- <5,00g <5,00g

/dev/sdc1 lvm2 --- <5,00g <5,00g

[exam@localhost ~]$ sudo pvdisplay

--- Physical volume ---

PV Name /dev/sda2

VG Name centos

PV Size <19,00 GiB / not usable 3,00 MiB

Allocatable yes (but full)

PE Size 4,00 MiB

Total PE 4863

Free PE 0

Allocated PE 4863

PV UUID KWvPQ8-6SqP-9nQo-XAGk-fWfs-vgEN-WHDj3H

"/dev/sdb1" is a new physical volume of "<5,00 GiB"

--- NEW Physical volume ---

PV Name /dev/sdb1

VG Name

PV Size <5,00 GiB

Allocatable NO

PE Size 0

Total PE 0

Free PE 0

Allocated PE 0

PV UUID hbcBI4-L9oN-icNx-oGAc-z4GA-3p44-8H7d2w

"/dev/sdc1" is a new physical volume of "<5,00 GiB"

--- NEW Physical volume ---

PV Name /dev/sdc1

VG Name

PV Size <5,00 GiB

Allocatable NO

PE Size 0

Total PE 0

Free PE 0

Allocated PE 0

PV UUID rHl2ed-Tw3Z-1isc-TtRu-eQTc-tYXy-gmbzoh

```

10. Создать две группы LVM и добавить в каждую из них по одному физическому тому из п.9.

```

[exam@localhost ~]$ sudo vgcreate vg1 /dev/sdb1

Volume group "vg1" successfully created

[exam@localhost ~]$ sudo vgcreate vg2 /dev/sdc1

Volume group "vg2" successfully created

[exam@localhost ~]$ sudo vgs

VG #PV #LV #SN Attr VSize VFree

centos 1 2 0 wz--n- <19,00g 0

vg1 1 0 0 wz--n- <5,00g <5,00g

vg2 1 0 0 wz--n- <5,00g <5,00g

```

11. В каждой из групп из п.10 создать логический том LVM размером 100% группы.

```

[exam@localhost ~]$ sudo lvcreate -n lv1 -l 100%FREE vg1

Logical volume "lv1" created.

[exam@localhost ~]$ sudo lvcreate -n lv2 -l 100%FREE vg2

Logical volume "lv2" created.

[exam@localhost ~]$ sudo lvs

LV VG Attr LSize Pool Origin Data% Meta% Move Log Cpy%Sync Convert

root centos -wi-ao---- <17,00g

swap centos -wi-ao---- 2,00g

lv1 vg1 -wi-a----- <5,00g

lv2 vg2 -wi-a----- <5,00g

```

12. На каждом логическом томе LVM создать файловую систему ext4.

```

[exam@localhost ~]$ sudo mkfs.ext4 /dev/vg1/lv1

mke2fs 1.42.9 (28-Dec-2013)

Filesystem label=

OS type: Linux

Block size=4096 (log=2)

Fragment size=4096 (log=2)

Stride=0 blocks, Stripe width=0 blocks

327680 inodes, 1309696 blocks

65484 blocks (5.00%) reserved for the super user

First data block=0

Maximum filesystem blocks=1342177280

40 block groups

32768 blocks per group, 32768 fragments per group

8192 inodes per group

Superblock backups stored on blocks:

32768, 98304, 163840, 229376, 294912, 819200, 884736

Allocating group tables: done

Writing inode tables: done

Creating journal (32768 blocks): done

Writing superblocks and filesystem accounting information: done

[exam@localhost ~]$ sudo mkfs.ext4 /dev/vg2/lv2

mke2fs 1.42.9 (28-Dec-2013)

Filesystem label=

OS type: Linux

Block size=4096 (log=2)

Fragment size=4096 (log=2)

Stride=0 blocks, Stripe width=0 blocks

327680 inodes, 1309696 blocks

65484 blocks (5.00%) reserved for the super user

First data block=0

Maximum filesystem blocks=1342177280

40 block groups

32768 blocks per group, 32768 fragments per group

8192 inodes per group

Superblock backups stored on blocks:

32768, 98304, 163840, 229376, 294912, 819200, 884736

Allocating group tables: done

Writing inode tables: done

Creating journal (32768 blocks): done

Writing superblocks and filesystem accounting information: done

```

13. Создать директории и использовать их в качестве точек монтирования файловых систем из п.12:

* /opt/mount1
* /opt/mount2

```

[exam@localhost ~]$ sudo mkdir /opt/mount{1..2}

```

14. Настроить систему так, чтобы монтирование происходило автоматически при запуске системы. Произвести монтирование новых файловых систем.

```

[root@localhost exam]# echo "/dev/vg1/lv1 /opt/mount1 ext4 defaults 0 0" >> /etc/fstab

[root@localhost exam]# echo "/dev/vg2/lv2 /opt/mount2 ext4 defaults 0 0" >> /etc/fstab

[exam@localhost ~]$ sudo mount -a

[exam@localhost ~]$ lsblk

NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT

sda 8:0 0 20G 0 disk

├─sda1 8:1 0 1G 0 part /boot

└─sda2 8:2 0 19G 0 part

├─centos-root 253:0 0 17G 0 lvm /

└─centos-swap 253:1 0 2G 0 lvm [SWAP]

sdb 8:16 0 5G 0 disk

└─sdb1 8:17 0 5G 0 part

└─vg1-lv1 253:2 0 5G 0 lvm /opt/mount1

sdc 8:32 0 5G 0 disk

└─sdc1 8:33 0 5G 0 part

└─vg2-lv2 253:3 0 5G 0 lvm /opt/mount2

sr0 11:0 1 1024M 0 rom

Finally CLONE VM1 as VM2 on VBOX

Internal Network VM2: enp0s8, IP: 192.168.2.12

```

Для **VM1** (шаги 15-16):

15. После монтирования создать 2 директории для хранения файлов Namenode сервиса HDFS:

* /opt/mount1/namenode-dir
* /opt/mount2/namenode-dir

```

[exam@localhost ~]$ sudo mkdir /opt/mount{1..2}/namenode-dir

```

16. Сделать пользователя **hdfs** и группу **hadoop** владельцами этих директорий.

```

[exam@localhost ~]$ sudo chown hdfs:hadoop /opt/mount{1..2}/namenode-dir

[exam@localhost ~]$ ll /opt/mount{1..2}

/opt/mount1:

total 20

drwx------. 2 root root 16384 Jan 21 15:15 lost+found

drwxr-xr-x. 2 hdfs hadoop 4096 Jan 21 17:07 namenode-dir

/opt/mount2:

total 20

drwx------. 2 root root 16384 Jan 21 15:15 lost+found

drwxr-xr-x. 2 hdfs hadoop 4096 Jan 21 17:07 namenode-dir

```

Для **VM2** (шаги 17-20):

17. После монтирования создать 2 директории для хранения файлов Datanode сервиса HDFS:

* /opt/mount1/datanode-dir
* /opt/mount2/datanode-dir

```

[exam@localhost ~]$ sudo mkdir /opt/mount{1..2}/datanode-dir

```

18. Сделать пользователя **hdfs** и группу **hadoop** владельцами директорий из п.17.

```

[exam@localhost ~]$ sudo chown hdfs:hadoop /opt/mount{1..2}/datanode-dir

```

19. Создать дополнительные 4 директории для Nodemanager сервиса YARN:

* /opt/mount1/nodemanager-local-dir
* /opt/mount2/nodemanager-local-dir
* /opt/mount1/nodemanager-log-dir
* /opt/mount2/nodemanager-log-dir

```

[exam@localhost ~]$ sudo mkdir /opt/mount{1..2}/nodemanager-local-dir

[exam@localhost ~]$ sudo mkdir /opt/mount{1..2}/nodemanager-log-dir

```

20. Сделать пользователя **yarn** и группу **hadoop** владельцами директорий из п.19.

```

[exam@localhost ~]$ sudo chown yarn:hadoop /opt/mount{1..2}/nodemanager\*

[exam@localhost ~]$ ll /opt/mount{1..2}

/opt/mount1:

total 28

drwxr-xr-x. 2 hdfs hadoop 4096 Jan 22 14:41 datanode-dir

drwx------. 2 root root 16384 Jan 21 15:15 lost+found

drwxr-xr-x. 2 yarn hadoop 4096 Jan 22 14:46 nodemanager-local-dir

drwxr-xr-x. 2 yarn hadoop 4096 Jan 22 14:46 nodemanager-log-dir

/opt/mount2:

total 28

drwxr-xr-x. 2 hdfs hadoop 4096 Jan 22 14:41 datanode-dir

drwx------. 2 root root 16384 Jan 21 15:15 lost+found

drwxr-xr-x. 2 yarn hadoop 4096 Jan 22 14:46 nodemanager-local-dir

drwxr-xr-x. 2 yarn hadoop 4096 Jan 22 14:46 nodemanager-log-dir

```

Для обеих машин:

21. Настроить доступ по SSH, используя ключи для пользователя **hadoop.**

**```**

[hadoop@vm1 exam]$ ssh-keygen

Generating public/private rsa key pair.

Enter file in which to save the key (/home/hadoop/.ssh/id\_rsa):

Created directory '/home/hadoop/.ssh'.

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

Your identification has been saved in /home/hadoop/.ssh/id\_rsa.

Your public key has been saved in /home/hadoop/.ssh/id\_rsa.pub.

The key fingerprint is:

SHA256:y2UuF+mZ1VORLrx+UYRUeLHfNF3aP+bxNdq9Gonj66Y hadoop@vm1

The key's randomart image is:

+---[RSA 2048]----+

| ..\*=|

| o+B|

| . o==|

| . + +\*|

| S = . ==\*|

| . \* =..\*oB|

| + \*o.+ o+|

| o.......|

| E++..o. |

+----[SHA256]-----+

[hadoop@vm2 exam]$ ssh-keygen

Generating public/private rsa key pair.

Enter file in which to save the key (/home/hadoop/.ssh/id\_rsa):

Created directory '/home/hadoop/.ssh'.

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

Your identification has been saved in /home/hadoop/.ssh/id\_rsa.

Your public key has been saved in /home/hadoop/.ssh/id\_rsa.pub.

The key fingerprint is:

SHA256:9CiuqLXc35Svrr+WJOGOZZ2SqXMwDOiEapIi88ChosQ hadoop@vm2

The key's randomart image is:

+---[RSA 2048]----+

| |

| |

|.. . |

|oo. .. o |

|B..o ..=S.. |

|@E +.O.+. |

|O+. O.+o. |

|.o.++.ooo. |

|..+ o+o=\*o. |

+----[SHA256]-----+

[hadoop@vm1 exam]$ ssh-copy-id hadoop@192.168.2.12

/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hadoop/.ssh/id\_rsa.pub"

The authenticity of host '192.168.2.12 (192.168.2.12)' can't be established.

ECDSA key fingerprint is SHA256:zEcOse55IYB33bYyZcebywXNL+QiVvPHU5E8lMJG4io.

ECDSA key fingerprint is MD5:02:95:06:94:e1:a8:3d:cf:00:a3:73:32:b6:63:f0:83.

Are you sure you want to continue connecting (yes/no)? yes

/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed

/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

hadoop@192.168.2.12's password:

Number of key(s) added: 1

Now try logging into the machine, with: "ssh 'hadoop@192.168.2.12'"

and check to make sure that only the key(s) you wanted were added.

[hadoop@vm2 exam]$ ssh-copy-id hadoop@192.168.2.11

/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hadoop/.ssh/id\_rsa.pub"

The authenticity of host '192.168.2.11 (192.168.2.11)' can't be established.

ECDSA key fingerprint is SHA256:zEcOse55IYB33bYyZcebywXNL+QiVvPHU5E8lMJG4io.

ECDSA key fingerprint is MD5:02:95:06:94:e1:a8:3d:cf:00:a3:73:32:b6:63:f0:83.

Are you sure you want to continue connecting (yes/no)? yes

/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed

/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys

hadoop@192.168.2.11's password:

Number of key(s) added: 1

Now try logging into the machine, with: "ssh 'hadoop@192.168.2.11'"

and check to make sure that only the key(s) you wanted were added.

[hadoop@vm1 exam]$ ssh hadoop@192.168.2.12

Last login: Sat Jan 22 17:01:17 2022

[hadoop@vm2 ~]$ exit

logout

Connection to 192.168.2.12 closed.

[hadoop@vm2 exam]$ ssh hadoop@192.168.2.11

Last login: Sat Jan 22 16:59:07 2022

[hadoop@vm1 ~]$ exit

logout

Connection to 192.168.2.11 closed.

**```**

22. Добавить VM1 и VM2 в /etc/hosts.

**```**

в /etc/hosts

192.168.2.11 VM1

192.168.2.12 VM2

[hadoop@vm1 exam]$ ssh hadoop@VM2

The authenticity of host 'vm2 (192.168.2.12)' can't be established.

ECDSA key fingerprint is SHA256:zEcOse55IYB33bYyZcebywXNL+QiVvPHU5E8lMJG4io.

ECDSA key fingerprint is MD5:02:95:06:94:e1:a8:3d:cf:00:a3:73:32:b6:63:f0:83.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'vm2' (ECDSA) to the list of known hosts.

Last login: Sat Jan 22 17:05:46 2022 from vm1

[hadoop@vm2 ~]$ exit

logout

Connection to vm2 closed.

[hadoop@vm2 exam]$ ssh hadoop@VM1

The authenticity of host 'vm1 (192.168.2.11)' can't be established.

ECDSA key fingerprint is SHA256:zEcOse55IYB33bYyZcebywXNL+QiVvPHU5E8lMJG4io.

ECDSA key fingerprint is MD5:02:95:06:94:e1:a8:3d:cf:00:a3:73:32:b6:63:f0:83.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'vm1' (ECDSA) to the list of known hosts.

Last login: Sat Jan 22 17:06:23 2022 from vm2

[hadoop@vm1 ~]$ exit

logout

Connection to vm1 closed.

**```**

23. Скачать файлы по ссылкам в /usr/local/hadoop/current/etc/hadoop/{hadoop-env.sh,core-site.xml,hdfs-site.xml,yarn-site.xml}. При помощи sed заменить заглушки на необходимые значения

* hadoop-env.sh (<https://gist.github.com/rdaadr/2f42f248f02aeda18105805493bb0e9b>)

Необходимо определить переменные **JAVA\_HOME** (путь до директории с OpenJDK8, установленную в п.3), **HADOOP\_HOME** (необходимо указать путь к симлинку из п.6) и **HADOOP\_HEAPSIZE\_MAX** (укажите значение в 512M)

* core-site.xml (<https://gist.github.com/rdaadr/64b9abd1700e15f04147ea48bc72b3c7>)

Необходимо указать имя хоста, на котором будет запущена HDFS Namenode (VM1)

* hdfs-site.xml (<https://gist.github.com/rdaadr/2bedf24fd2721bad276e416b57d63e38>)

Необходимо указать директории namenode-dir, а также datanode-dir, каждый раз через запятую (например, /opt/mount1/namenode-dir,/opt/mount2/namenode-dir)

* yarn-site.xml (<https://gist.github.com/Stupnikov-NA/ba87c0072cd51aa85c9ee6334cc99158>)

Необходимо подставить имя хоста, на котором будет развернут YARN Resource Manager (VM1), а также пути до директорий nodemanager-local-dir и nodemanager-log-dir (если необходимо указать несколько директорий, то необходимо их разделить запятыми)

```

[hadoop@vm1 exam]$ cd /usr/local/hadoop/current/etc/hadoop

[hadoop@vm1 hadoop]$ curl -O https://gist.githubusercontent.com/rdaadr/2f42f248f02aeda18105805493bb0e9b/raw/6303e424373b3459bcf3720b253c01373666fe7c/hadoop-env.sh

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 15980 100 15980 0 0 24547 0 --:--:-- --:--:-- --:--:-- 24546

[hadoop@vm1 hadoop]$ curl -O https://gist.githubusercontent.com/rdaadr/64b9abd1700e15f04147ea48bc72b3c7/raw/2d416bf137cba81b107508153621ee548e2c877d/core-site.xml

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 908 100 908 0 0 1397 0 --:--:-- --:--:-- --:--:-- 1399

[hadoop@vm1 hadoop]$ curl -O https://gist.githubusercontent.com/rdaadr/2bedf24fd2721bad276e416b57d63e38/raw/640ee95adafa31a70869b54767104b826964af48/hdfs-site.xml

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 1081 100 1081 0 0 1823 0 --:--:-- --:--:-- --:--:-- 1822

[hadoop@vm1 hadoop]$ curl -O https://gist.githubusercontent.com/Stupnikov-NA/ba87c0072cd51aa85c9ee6334cc99158/raw/bda0f760878d97213196d634be9b53a089e796ea/yarn-site.xml

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 1499 100 1499 0 0 4590 0 --:--:-- --:--:-- --:--:-- 4584

[hadoop@vm1 hadoop]$ sed -i 's|"%PATH\_TO\_OPENJDK8\_INSTALLATION%"|/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.312.b07-1.el7\_9.x86\_64/jre|' hadoop-env.sh

[hadoop@vm1 hadoop]$ sed -i 's|"%PATH\_TO\_HADOOP\_INSTALLATION"|/usr/local/hadoop/current|' hadoop-env.sh

[hadoop@vm1 hadoop]$ sed -i 's|"%HADOOP\_HEAP\_SIZE%"|512m|' hadoop-env.sh

[hadoop@vm1 hadoop]$ sed -i 's|%HDFS\_NAMENODE\_HOSTNAME%|VM1|' core-site.xml

[hadoop@vm1 hadoop]$ sed -i 's|%YARN\_RESOURCE\_MANAGER\_HOSTNAME%|VM1|' yarn-site.xml

(на VM2 тоже)

Для VM1

[hadoop@vm1 hadoop]$ sed -i 's|%NAMENODE\_DIRS%|/opt/mount1/namenode-dir,/opt/mount2/namenode-dir|' hdfs-site.xml

Для VM2

[hadoop@vm2 hadoop]$ sed -i 's|%DATANODE\_DIRS%|/opt/mount1/datanode-dir,/opt/mount2/datanode-dir|' hdfs-site.xml

[hadoop@vm2 hadoop]$ sed -i 's|%NODE\_MANAGER\_LOCAL\_DIR%|/opt/mount1/nodemanager-local-dir,/opt/mount2/nodemanager-local-dir|' yarn-site.xml

[hadoop@vm2 hadoop]$ sed -i 's|%NODE\_MANAGER\_LOG\_DIR%|/opt/mount1/nodemanager-log-dir,/opt/mount2/nodemanager-log-dir|' yarn-site.xml

```

24. Задать переменную окружения **HADOOP\_HOME** через /etc/profile

```

sudo vi /etc/profile

HADOOP\_HOME=/usr/local/hadoop/current

export HADOOP\_HOME

[exam@vm1 ~]$ source /etc/profile

[exam@vm1 ~]$ echo $HADOOP\_HOME

/usr/local/hadoop/current

```

SOME FIREWALL RULES:

```

firewall-cmd --permanent --add-port=9870/tcp

firewall-cmd --permanent --add-port=8088/tcp

firewall-cmd --permanent --add-port=9866/tcp

firewall-cmd --permanent --add-port=9864/tcp

firewall-cmd --permanent --add-port=8020/tcp

firewall-cmd --permanent --add-port=8031/tcp

firewall-cmd --permanent --add-port=8042/tcp

firewall-cmd --reload

```

Для **VM1** (шаги 25-26):

25. Произвести форматирование HDFS (от имени пользователя **hdfs**):

* $HADOOP\_HOME/bin/hdfs namenode -format cluster1

```

[hadoop@vm1 hadoop]$ mkdir /usr/local/hadoop/current/logs

(hdfs не мог создать эту папку)

[hdfs@vm1 ~]$ $HADOOP\_HOME/bin/hdfs namenode -format cluster1

2022-01-22 21:22:38,527 INFO namenode.NameNode: STARTUP\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

STARTUP\_MSG: Starting NameNode

STARTUP\_MSG: host = VM1/192.168.2.11

STARTUP\_MSG: args = [-format, cluster1]

STARTUP\_MSG: version = 3.1.2

...

...

2022-01-22 21:22:43,054 INFO namenode.NameNode: SHUTDOWN\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SHUTDOWN\_MSG: Shutting down NameNode at VM1/192.168.2.11

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

```

26. Запустить демоны сервисов Hadoop:

Для запуска Namenode (от имени пользователя **hdfs**):

* $HADOOP\_HOME/bin/hdfs --daemon start namenode

Для запуска Resource Manager (от имени пользователя **yarn**):

* $HADOOP\_HOME/bin/yarn --daemon start resourcemanager

```

[hdfs@vm1 ~]$ $HADOOP\_HOME/bin/hdfs --daemon start namenode

[yarn@vm1 exam]$ $HADOOP\_HOME/bin/yarn --daemon start resourcemanager

```

Для **VM2** (шаг 27):

27. Запустить демоны сервисов:

Для запуска Datanode (от имени **hdfs**):

* $HADOOP\_HOME/bin/hdfs --daemon start datanode

Для запуска Node Manager (от имени **yarn**):

* $HADOOP\_HOME/bin/yarn --daemon start nodemanager

```

[hdfs@vm2 ~]$ $HADOOP\_HOME/bin/hdfs --daemon start datanode

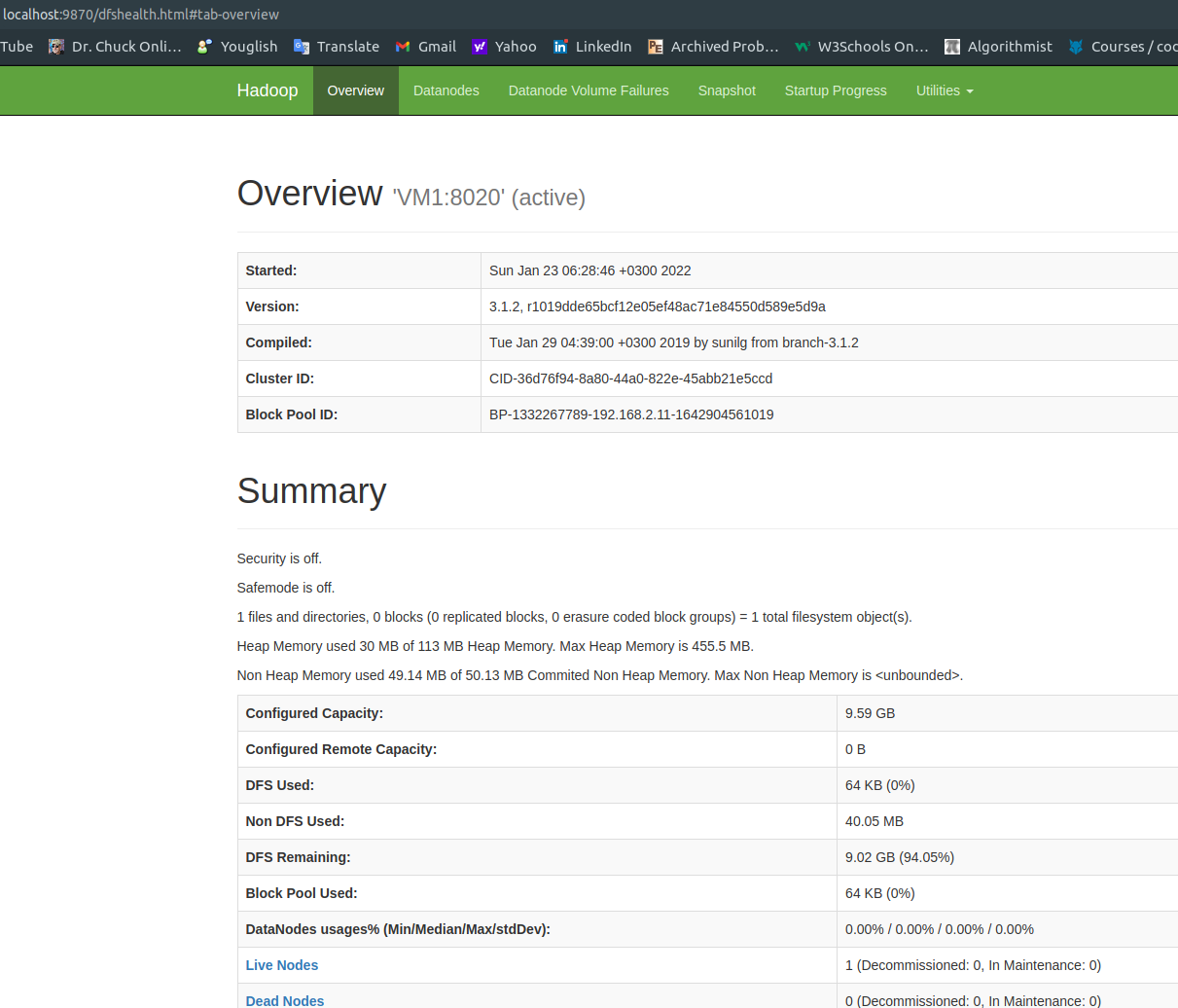
[yarn@vm2 hdfs]$ $HADOOP\_HOME/bin/yarn --daemon start nodemanager

```

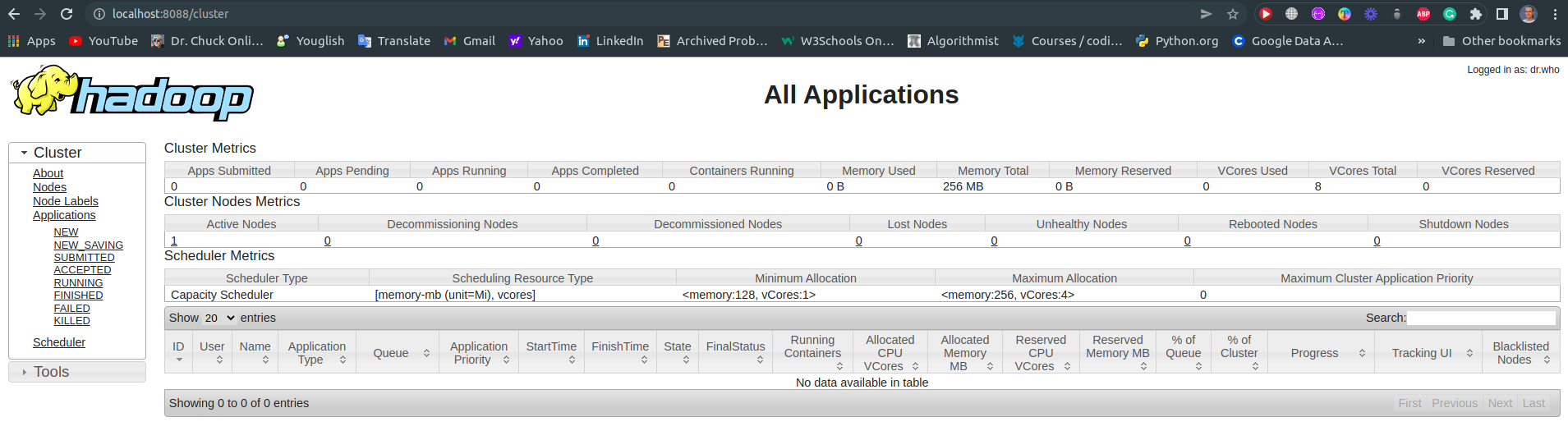
28. Проверить доступность Web-интефейсов HDFS Namenode и YARN Resource Manager по портам 9870 и 8088 соответственно (VM1). << порты должны быть доступны с хостовой системы.

```

ssh exam@127.0.0.1 -p 2223 -L 9870:192.168.2.11:9870



ssh exam@127.0.0.1 -p 2223 -L 8088:192.168.2.11:8088



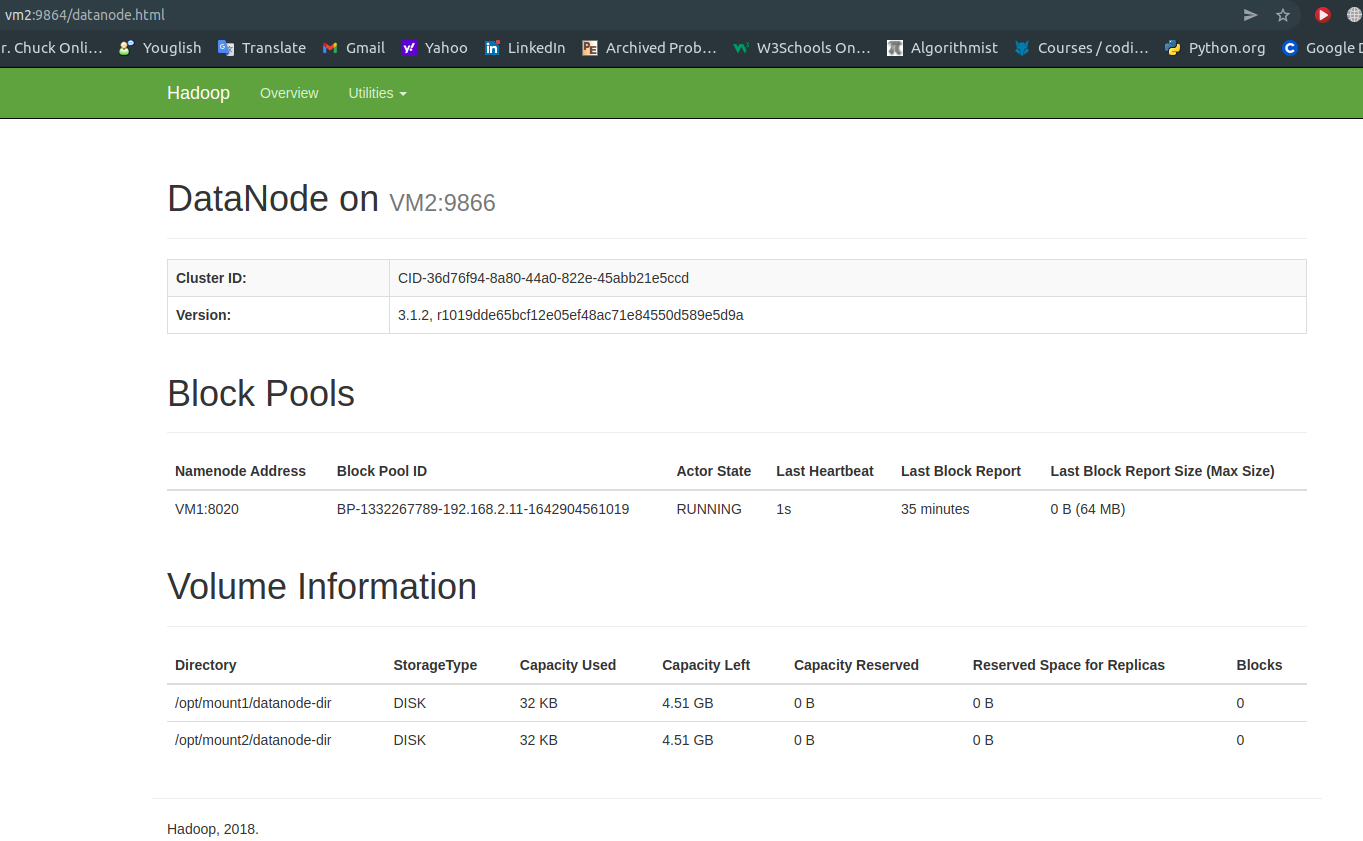
На хосте

```

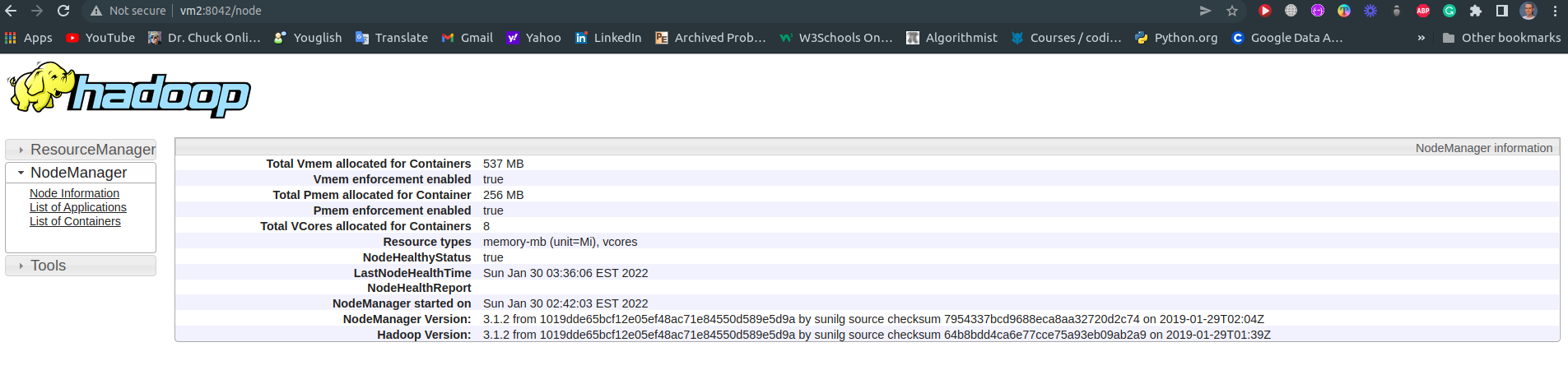
127.0.0.1 localhost vm2 vm1

```

ssh exam@127.0.0.1 -p 2224 -L 9864:192.168.2.12:9864



ssh exam@127.0.0.1 -p 2224 -L 8042:192.168.2.12:8042



```

29. Настроить управление запуском каждого компонента Hadoop при помощи systemd (используя юниты-сервисы).

```

VM1

"/etc/systemd/system/namenode.service"

[Unit]

Description=Hadoop namenode

[Service]

User=hdfs

Group=hadoop

Type=forking

ExecStart=/usr/local/hadoop/current/bin/hdfs --daemon start namenode

ExecStop=/usr/local/hadoop/current/bin/hdfs --daemon stop namenode

WorkingDirectory=/usr/local/hadoop

Environment=JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.312.b07-1.el7\_9.x86\_64/jre

Environment=HADOOP\_HOME=/usr/local/hadoop/current

[Install]

WantedBy=multi-user.target

"/etc/systemd/system/resourcemanager.service"

[Unit]

Description=Hadoop resourcemanager

[Service]

User=yarn

Group=hadoop

Type=forking

ExecStart=/usr/local/hadoop/current/bin/yarn --daemon start resourcemanager

ExecStop=/usr/local/hadoop/current/bin/yarn --daemon stop resourcemanager

WorkingDirectory=/usr/local/hadoop

Environment=JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.312.b07-1.el7\_9.x86\_64/jre

Environment=HADOOP\_HOME=/usr/local/hadoop/current

[Install]

WantedBy=multi-user.target

VM2

"/etc/systemd/system/datanode.service"

[Unit]

Description=Hadoop datanode

[Service]

User=hdfs

Group=hadoop

Type=forking

ExecStart=/usr/local/hadoop/current/bin/hdfs --daemon start datanode

ExecStop=/usr/local/hadoop/current/bin/hdfs --daemon stop datanode

WorkingDirectory=/usr/local/hadoop

Environment=JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.312.b07-1.el7\_9.x86\_64/jre

Environment=HADOOP\_HOME=/usr/local/hadoop/current

[Install]

WantedBy=multi-user.target

"/etc/systemd/system/nodemanager.service"

[Unit]

Description=Hadoop nodemanager

[Service]

User=yarn

Group=hadoop

Type=forking

ExecStart=/usr/local/hadoop/current/bin/yarn --daemon start nodemanager

ExecStop=/usr/local/hadoop/current/bin/yarn --daemon stop nodemanager

WorkingDirectory=/usr/local/hadoop

Environment=JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.312.b07-1.el7\_9.x86\_64/jre

Environment=HADOOP\_HOME=/usr/local/hadoop/current

[Install]

WantedBy=multi-user.target

```

**Полезные ссылки:**

* <https://hadoop.apache.org/docs/r3.1.2/hadoop-project-dist/hadoop-common/ClusterSetup.html>