

Домашнее задание №1

БПИ227 Артемьев Александр

Следующая инструкция представляет из себя пошаговое руководство по развертыванию кластера hdfs, включающего в себя 3 DataNode и обязательные для функционирования кластера сервисы: NameNode и Secondary NameNode.

Развертывание кластера происходит на 3 узлах + узел edge-node. На первом узле будут расположены NameNode, Secondary NameNode и DataNode(узел nn). Также еще два узла на каждом из которых будет по одной DataNode(узлы dn-00 и dn-01). И узел edge-node (узел jn). Собственно, для развертывания у нас должны быть их ip, далее в работе обозначим ip каждой ноды соответственно следующим образом:

jn_glob_ip – узел для входа в jn
jn_local_ip – локальный адрес jn
nn_ip – локальный адрес nn
dn-00_ip – локальный адрес dn-00
dn-01_ip – локальный адрес dn-01

Также для подключения нам необходимо имя пользователя, который может подключиться к jn(user_name). Далее в командах, <param> означает значение переменной param.

Порядок действий:

1) Подключаемся к edge-node:

```
ssh <user_name>@<jn_glob_ip>
```

2) Для того, чтобы переходить по нодам не используя пароль, будем использовать ssh. Генерируем ssh ключ:

```
ssh-keygen
```

3) Добавляем созданный ключ в число авторизованных:

```
cat .ssh/id_ed25519.pub >> .ssh/authorized_keys
```

4) Теперь скопируем созданный ключ по всем узлам:

```
scp .ssh/authorized_keys <nn_ip>:/home/<user_name>/.ssh
```

```
scp .ssh/authorized_keys <dn-00_ip>:/home/<user_name>/.ssh
```

```
scp .ssh/authorized_keys <dn-01_ip>:/home/<user_name>/.ssh
```

5) Следующим шагом мы хотим настроить возможность обращаться внутри нод к другим нодам по имени, для этого в каждой ноде поменяем /etc/hosts. Также создадим пользователя на каждой ноде от имени которого будут выполняться сервисы Hadoop. Для этого заходим в jn, nn, dn-00, dn-01 и делаем следующее:

`sudo adduser hadoop` (здесь надо будет ввести пароль для созданного пользователя)

`sudo vim /etc/hosts`

Находясь в файле `/etc/hosts` комментируем все строки и добавляем 4 строки:

Для jn:

#..

#..

127.0.0.1 jn

<nn_ip> nn

<dn-00_ip> dn-00

<dn-01_ip> dn-01

#..

#..

Для nn:

#..

#..

<jn_local_ip> jn

<nn_ip> nn

<dn-00_ip> dn-00

<dn-01_ip> dn-01

#..

#..

Для dn-00:

#..

#..

<jn_local_ip> jn

<nn_ip> nn

127.0.0.1 dn-00

<dn-01_ip> dn-01

#..

#..

для dn-01:

#..

#..

<jn_local_ip> jn

<nn_ip> nn

<dn-00_ip> dn-00

127.0.0.1 dn-01

#..

#..

Затем меняем в текущей node в файле `/etc/hostname` имя на указанное в `/etc/hosts`(соответственно jn, nn,dn-00, dn-01)

`sudo vim /etc/hostname`

б)К этому шагу мы поменяли в каждой ноде по 2 файла `/etc/hostname` и `/etc/hosts`, а также в каждой ноде добавили пользователя `hadoop`. Теперь переходим в пользователя `hadoop`:

```
sudo -i -u hadoop
```

7) И скачаем дистрибутив с hadoop:

```
wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
```

8) Теперь сделаем те же манипуляции с генерацией ключа ssh и его распространении на остальные ноды:

```
ssh-keygen
```

```
cat .ssh/id_ed25519.pub >> .ssh/authorized_keys
```

```
scp -r .ssh/ nn:/home/hadoop
```

```
scp -r .ssh/ dn-00:/home/hadoop
```

```
scp -r .ssh/ dn-01:/home/hadoop
```

9)Скопируем установленный дистрибутив на остальные ноды:

```
scp hadoop-3.4.0.tar.gz nn:/home/hadoop
```

```
scp hadoop-3.4.0.tar.gz dn-00:/home/hadoop
```

```
scp hadoop-3.4.0.tar.gz dn-01:/home/hadoop
```

10) Распаковываем архив с дистрибутивом на каждой ноде(переход между нодами через ssh <название ноды nn, jn, dn-01,dn-00 куда хотим перейти>):

```
tar -xzf hadoop-3.4.0.tar.gz
```

11)Убеждаемся, что среда в которой мы работаем соответствует требованию hadoop, а именно проверяем, что у нас установлена java подходящей версии:

```
java -version
```

12)Теперь находим директорию, где находится java

```
which java – (путь до бинарника)
```

```
readlink -f <путь до бинарника из предыдущего шага> (запомним этот путь как home_dir_java)
```

13) Настраиваем наш профиль

```
vim .profile
```

(Добавляем следующие строки:

```
export HADOOP_HOME=/home/hadoop/hadoop-3.4.0
```

```
export JAVA_HOME= <../home_dir_java>
```

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

)

(затем активируем изменения)

```
source .profile
```

14) Теперь раскидываем конфиг в profile во все ноды

```
scp .profile nn:/home/hadoop
```

```
scp .profile dn-00:/home/hadoop
```

```
scp .profile dn-01:/home/hadoop
```

15) Следующий шаг изменить конфиги в jn

```
cd hadoop-3.4.0/etc/hadoop
```

```
vim hadoop-env.sh
```

Добавляем путь к java `JAVA_HOME= <../../home_dir_java>`

```
vim core-site.xml
```

меняем конфиг на:

```
<configuration>
```

```
<property>
```

```
  <name>fs.defaultFS</name>
```

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

```
</property>
```

```
</configuration>
```

```
vim hdfs-site.xml
```

меняем конфиг на:

```
<configuration>
```

```
<property>
```

```
  <name>dfs.replication</name>
```

```
  <value>3 </value>
```

```
</property>
```

```
</configuration>
```

Последнее что мы хотим изменить – список workers(список node, где будет запущен DataNode)

```
vim workers
```

Перезаписываем на:

```
nn
```

```
dn-00
```

```
dn-01
```

16) Затем мы скопируем конфиги на оставшиеся узлы:

```
scp hadoop-env.sh nn:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp hadoop-env.sh dn-00:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp hadoop-env.sh dn-01:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp core-site.xml nn:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp core-site.xml dn-00:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp core-site.xml dn-01:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp hdfs-site.xml nn:/home/hadoop/hadoop-3.4.0/etc/hadoop
scp hdfs-site.xml dn-00:/home/hadoop/hadoop-3.4.0/etc/hadoop
scp hdfs-site.xml dn-01:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

```
scp workers nn:/home/hadoop/hadoop-3.4.0/etc/hadoop
scp workers dn-00:/home/hadoop/hadoop-3.4.0/etc/hadoop
scp workers dn-01:/home/hadoop/hadoop-3.4.0/etc/hadoop
```

17) Подключаемся к nn:

```
ssh nn
```

18) Отформатируем и создадим файловую систему

```
cd hadoop-3.4.0/
bin/hdfs namenode -format
```

19) Запускаем hdfs

```
sbin/start-dfs.sh
```

Мои данные:

Env team-1

User team

пароль для входа х=T35T_sMdm4

узел для входа 176.109.91.3

jn 192.168.1.6

nn 192.168.1.7

dn-00 192.168.1.8

dn-01 192.168.1.9