

Домашнее задание №1

Выполнил: **Умертаев Арслан Наушанович БПИ227**

Исходные данные:

Узел для входа **176.109.91.11** - <global_ip>, JumpNode 192.168.1.38 - <local_ip>,

LocalNodeIp: NameNode 192.168.1.39, DataNode-00 192.168.1.40, DataNode-01 192.168.1.41

Username: team

Имеющиеся ноды **team-9-jn, team-9-nn, team-9-dn-00, team-9-dn-01**

Инициализация

1. Открываем терминал подключаемся к JumpNode, введя ssh **team@176.109.91.11**
ssh <user_name>@<global_host_ip> -- общий случай подключения
 - 1.1 **ssh-keygen** # генерация ssh ключа
 - 1.2 **cat .ssh/id_ed25519.pub >> .ssh/authorized_keys** # добавление ключа в число авторизованных
2. Распространяем сгенерированный ключ по всем нодам **scp .ssh/authorized_keys <ip_node>:/home/<user_name>/.ssh/**
3.
 1. Для того, чтобы обращаться по именам к нодам нужно поменять файл /etc/hosts в каждой ноде, закомментировав соответствующую строку (строку узла, на котором находимся в текущий момент).

```
127.0.0.1 team-9-jn # не меняется
#Строки, которые нужно по одной закомментировать
192.168.1.30    team-9-jn # JumpNode
192.168.1.31    team-9-nn # NameNode
192.168.1.32    team-9-dn-00 # DataNode 00
192.168.1.33    team-9-dn-01 # DataNode 01

#Важно для nn в этом файле должны быть только
192.168.1.30    team-9-jn # JumpNode
192.168.1.32    team-9-dn-00 # DataNode 00
192.168.1.33    team-9-dn-01 # DataNode 01
```

2. Заменить текст в **/etc/hostname** на соответствующие директории ноды **team-9-jn, team-9-nn, team-9-dn-00, team-9-dn-01**
3. Теперь все ноды могут обращаться друг к другу по имени
4. Для каждой node создаем пользователя hadoop, с помощью команды **sudo add user hadoop**
Теперь ко всем нодам можем обращаться по имени
5. Скачать дистрибутив с hadoop **wget <https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz>**
6. Переключаемся на пользователя hadoop: **sudo -i -u hadoop**
7. Теперь нужно так, чтобы пользователь Hadoop заходил на узлы от своего имени
 1. Генерируем ssh ключ для hadoop (повторяем п 1.1)
 2. Добавляем в список авторизованных ключей (повторяем п 1.2)
 3. Распространяем на остальные ноды: **scp -r .ssh/<node_name>:/home/hadoop**
8. После того, как скачался дистрибутив с hadoop его также нужно распространить на все ноды
 1. Копируем на все ноды **scp hadoop-3.4.0.tar.gz team-9-<node_name>:/home/hadoop**
 2. Подключаемся к каждой ноде и распаковываем архивы на всех нодах: **tar -xzvf hadoop-3.4.0.tar.gz**

9. Настроить Java, находясь в hadoop jn

1. Проверить наличие Java, если нет то установить. **java -version, which java**
2. Найти путь до Java и скопировать его **readlink -f <результат which java>**
3. Добавить в файл .profile

```
export HADOOP_HOME=/home/hadoop/hadoop-3.4.0
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

4. **source .profile** - активируем изменения (проверить можно с помощью hadoop version)
5. Раскидать этот файл на все ноды: **scp .profile team-9-<name_node>:/home/hadoop**
6. На jn в папке ~/hadoop-3.4.0/etc/hadoop/ поправить с помощью vim :
 1. файл hadoop-env.sh (добавить **JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64**)
 2. файл core-site.xml

```
<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://team-9-nn:9000</value>
</property>
</configuration>
```

3. файл hdfs-site.xml

```
<configuration>
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
</configuration>
```

4. Workers

```
team-9-nn  
team-9-dn-00  
team-9-dn-01
```

7. Копируем все файлы выше на все ноды **scp <file_name> team-9-<name_node>:/home/hadoop/hadoop-3.4.0/etc/hadoop**
8. Переходим на **namenode**: **ssh team-9-nn** под пользователем **hadoop**
9. Отформатировать namenode: **hadoop-3.4.0/bin/hdfs namenode -format**
10. Запустить файловую систему: **hadoop-3.4.0/sbin/start-dfs.sh**
11. Проверим, что все запущено, выполнив команду команду **jps**. Должно вывестись: **DataNode**, **NameNode** и **SecondaryNameNode**.