

Инструкция по развертыванию Apache Hive

Предполагается что кластер Hadoop и Yarn уже запущены

Мои данные:

Узел для входа 176.109.91.28

Jump node 192.168.1.118

Name node 192.168.1.119

Data node-00 192.168.1.120

Data node-01 192.168.1.121

ssh team@176.109.91.28 -- заходим на jump-ноду

ssh nn

sudo apt install postgresql -- устанавливаем postgres

sudo -i -u postgres -- меняем пользователя на postgres

psql

CREATE DATABASE metastore; -- создаем базу данных с названием metastore

CREATE USER hive with password 'hiveMegaPass'; -- создаем пользователя hive с паролем hiveMegaPass

GRANT ALL PRIVILEGES ON DATABASE "metastore" to hive; -- даем пользователю hive все привилегии на базе metastore

ALTER DATABASE metastore OWNER TO hive; -- передаем владение базой данных metastore пользователю hive

\q

exit -- должны оказаться на name ноде от лица пользователя team

sudo vim /etc/postgresql/16/main/postgresql.conf

Устанавливаем listen_addresses = '*':

```
listen_addresses = '*'
```

Сохраняем файл и выходим из редактора

sudo vim /etc/postgresql/16/main/pg_hba.conf

Добавляем в файл следующие строчки:

host	metastore	hive	192.168.1.1/32	password
host	metastore	hive	192.168.1.118/32	password – узел джамп ноды

Сохраняем файл и выходим из редактора

sudo systemctl restart postgresql -- перезапускаем postgres

exit -- должны оказаться на jump ноде от имени пользователя team

```
sudo apt install postgresql-client-16
```

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

```
wget https://archive.apache.org/dist/hive/hive-4.0.0-alpha-2/apache-hive-4.0.0-alpha-2-bin.tar.gz
```

```
tar -xzf apache-hive-4.0.0-alpha-2-bin.tar.gz
```

```
cd apache-hive-4.0.0-alpha-2-bin/lib/
```

```
wget https://jdbc.postgresql.org/download/postgresql-42.7.4.jar
```

```
vim ../conf/hive-site.xml
```

Добавляем в файл следующее содержание:

```
<configuration>
  <property>
    <name>hive.server2.authentication</name>
    <value>NONE</value>
  </property>
  <property>
    <name>hive.metastore.warehouse.dir</name>
    <value>/user/hive/warehouse</value>
  </property>
  <property>
    <name>hive.server2.thrift.port</name>
    <value>5433</value>
    <description>TCP port number to listenon, default 10000</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:postgresql://nn:5432/metastore</value>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>org.postgresql.Driver</value>
  </property>
  <property>
```

```
<name>javax.jdo.option.ConnectionUserName</name>
<value>hive</value>
</property>
<property>
<name>javax.jdo.option.ConnectionPassword</name>
<value>hiveMegaPass</value>
</property>
</configuration>
```

Сохраняем файл и выходим из редактора

vim ~/.profile

Добавляем в файл следующие строки:

```
export HIVE_HOME=/home/hadoop/apache-hive-4.0.0-alpha-2-bin
export HIVE_CONF_DIR=$HIVE_HOME/conf
export HIVE_AUX_JARS_PATH=$HIVE_HOME/lib/*
export PATH=$PATH:$HIVE_HOME/bin
```

Сохраняем файл и выходим из редактора

source ~/.profile

hdfs dfs -mkdir -p /user/hive/warehouse

hdfs dfs -chmod g+w /tmp

hdfs dfs -chmod g+w /user/hive/warehouse

cd ..

bin/schematool -dbType postgres -initSchema

tmux -- для безопасности запустим hive в сессионном менеджере

**hive --hiveconf hive.server2.enable.doAs=false --hiveconf
hive.security.authorization.enabled=false --service hiveserver2 1>> /tmp/hs2.log 2>>
/tmp/hs2.log &**

Выходим из сессии tmux (Ctrl + b d)

beeline -u jdbc:hive2://jn:5433 -n scott -p tiger

!q

Попробуем создать таблицу в базе данных:

hdfs dfs -put colours.tsv /test

beeline -u jdbc:hive2://jn:5433 -n scott -p tiger

CREATE DATABASE test;

CREATE TABLE IF NOT EXISTS test.colours (

colour string,

number int)

COMMENT 'colours table'

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';

Загружаем данные:

LOAD DATA INPATH '/test/colours.tsv' INTO TABLE test.colours;