CONFIGURAZIONE CLUSTER

Simone Cullino *Roger Ferrod*

CentOS 7 minimal 1810 x86_64

Per semplicità e risparmio di risorse, i dati non sono replicati nel cluster
Inoltre il firewall è disabilitato in ogni macchina

Nodi collegati in rete locale tramite ethernet
Configurazione IP statico tramite DHCP

CONFIGURAZIONE MACCHINE

```
yum update
yum install nano
yum install net-tools

nano /etc/sysconfig/selinux
      SELINUX=disabled

systemctl disable firewalld

systemctl stop firewalld


nano /etc/hostname
      <hostname>

service network restart

nano /etc/hosts
      <ip addr>  master
      <ip addr>  shard1
      <ip addr>  shard2
      ...         ....

## hosts Windows
C:\Windows\System32\drivers\etc\hosts
      <ip addr>  master
      <ip addr>  shard1
      <ip addr>  shard2
      ...         ....
```

MONGODB

```
nano /etc/yum.repos.d/mongodb-org-4.0.repo
      [mongodb-org-4.0]
      name=MongoDB Repository
      baseurl=https://repo.mongodb.org/yum/redhat/$releasever/mongodb-
org/4.0/x86_64/
      gpgcheck=0
      enabled=1
      gpgkey=https://www.mongodb.org/static/pgp/server-4.0.asc

yum update
yum install mongodb-org
mongod --version
```

REPLICA SET CONFIG

```
## avviare i server config con:
mongod --configsvr --replSet repl1 --dbpath /var/lib/mongo --bind_ip 0.0.0.0 --
port 27019

## da shell client
mongo --host <primary_config_hostname> --port 27019
rs.initiate(
  {
    _id : "repl1",
    members: [
      { _id : 0, host : "master:27019" }
    ]
  }
)

rs.status()
```

SHARDING

```
## da shell client
mongo --host <router hostname> --port 27017

sh.addShard("<shard>:27018")
sh.status()
```

SERVICE

server config

```
nano /etc/systemd/system/mongoconfig.service
      [Unit]
      Description=MongoDB server

      [Service]
      ExecStart=/usr/bin/mongod --configsvr --replSet repl1 --dbpath
      /var/lib/mongo --bind_ip 0.0.0.0 --port 27019

      [Install]
      WantedBy=multi-user.target

systemctl enable mongoconfig
```

router

```
nano /etc/systemd/system/mongorouter.service
      [Unit]
      Description=MongoDB server

      [Service]
      ExecStart=/usr/bin/mongos --configdb repl1/master:27019 --bind_ip 0.0.0.0
      --port 27017

      [Install]
      WantedBy=multi-user.target

systemctl enable mongorouter
```

shard

```
nano /etc/systemd/system/mongoshard.service
      [Unit]
      Description=MongoDB server

      [Service]
      ExecStart=/usr/bin/mongod --shardsvr --dbpath /var/lib/mongo --bind_ip
      0.0.0.0 --port 27018

      [Install]
      WantedBy=multi-user.target

systemctl enable mongoshard
```

SPARK

```
yum install java-1.8.0-openjdk

cd /opt

wget http://www-eu.apache.org/dist/spark/spark-2.4.2/spark-2.4.2-bin-hadoop2.7.tgz

tar -xzf spark-2.4.2-bin-hadoop2.7.tgz

ln -s /opt/spark-2.4.2-bin-hadoop2.7 /opt/spark


export PATH=$PATH:/opt/spark/bin

export SPARK_HOME=/opt/spark

export JAVA_HOME=/usr/lib/jvm/jre-1.8.0-openjdk

export PATH=$PATH:$JAVA_HOME


cp $SPARK_HOME/conf/spark-env.sh.template $SPARK_HOME/conf/spark-env.sh

echo 'SPARK_MASTER_HOST=master' >> $SPARK_HOME/conf/spark-env.sh
```

SSH

```
## slaves

mkdir ~/.ssh


## master

ssh-keygen -t rsa


cat ~/.ssh/id_rsa.pub | ssh root@shard1 'cat >> ~/.ssh/authorized_keys'

…

echo -e "shard1\nshard2\nshard3" > $SPARK_HOME/conf/slaves
```

POSTGRES

```
yum install https://download.postgresql.org/pub/repos/yum/9.6/redhat/rhel-7-x86_64/pgdg-redhat96-9.6-3.noarch.rpm
yum install postgresql96-server

/usr/pgsql-9.6/bin/postgresql96-setup initdb
systemctl enable postgresql-9.6.service
systemctl start postgresql-9.6.service

su – postgres
psql
\password postgres
<password>
\q
```

NETWORK CONFIGURATION

```
nano /var/lib/pgsql/9.6/data/pg_hba.conf
      # "local" is for Unix domain socket connections only
      local   all all                                     md5
      # IPv4 local connections:
      host    all     all           192.168.235.0/24      md5
      # IPv6 local connections:
      host    all     all           ::1/128               md5

nano /var/lib/pgsql/9.6/data/postgresql.conf
      listen_addresses = '*'

systemctl restart postgresql-9.6
```

START

- MongoDB

```
## shard
systemctl start mongoshard
## master
systemctl start mongoconfig
systemctl start mongorouter
```

- Postgres

```
## master
systemctl start postgresql-9.6
```

- Spark

```
## master
$SPARK_HOME/sbin/start-all.sh
```

SUBMIT & RUNNING

distribuire jar in ogni nodo del cluster

spark-submit --class *<mainClass>* --master spark://*<master>*:7077 --deploy-mode cluster file:*<local_jar>* *<params…>*

spark-submit --driver-class-path <jdbc_jar> --class <mainClass> --master spark://<master>:7077 --deploy-mode cluster file: *<local_jar>* *<params…>*

java -jar <jar_file> <params…>