Name: Chaudhari Ganesh Dadabhau

Configure Hadoop cluster using Ansible

Task No. - 4

1. Installation of Hadoop Requirements
2. Configuration of Name Node & Data Node
3. Starting Hadoop Services

lets start, **Ansible**

Ansible is configuration management tool. It works on push mechanism and it is agentless. ansible is built on the top of python hence before ansible installation we should have installed python3. we can install ansible using *pip3 install ansible*. after that configure /etc/ansible/ansible.cfg and inventory files like below

```
[root@localhost ~]# ansible --version
ansible 2.10.3
  config file = /etc/ansible/ansible.cfg
  configured module search path = ['/root/.ansible/plugins/modules', '/usr/share
/ansible/plugins/modules']
  ansible python module location = /usr/local/lib/python3.6/site-packages/ansibl
e
  executable location = /usr/local/bin/ansible
  python version = 3.6.8 (default, Jan 11 2019, 02:17:16) [GCC 8.2.1 20180905 (R
ed Hat 8.2.1-3)]
[root@localhost ~]# cat /etc/ansible/ansible.cfg
[defaults]
inventory= /etc/hosts.txt
host_key_checking = false
```

```
[root@localhost ~]# cat /etc/hosts.txt
[master]
192.168.225.203 ansible_user=root ansible_ssh_pass=root ansible_connection=ssh
[slave]
192.168.225.204 ansible_user=root ansible_ssh_pass=root ansible_connection=ssh
192.168.225.132 ansible_user=root ansible_ssh_pass=root ansible_connection=ssh
[root@localhost ~]#
```

lets write playbook for master nodes

1.Transfer java JDK and install it on target node because Hadoop built from java language then transfer Hadoop library which will be compactiable with java.

```
- hosts: master
  vars_files:
            - vars.yml
  tasks:
  - name: "Namenode copying file of hotspot java"
    copy:
        src: "/jdk-8u171-linux-x64.rpm"
        dest: "/"


  - name: "Namenode checking java is present or not"
    command:
             "jps"
    register: x
    ignore_errors: true

  - name: "Namenode installing java"
    command:
             "rpm -i /jdk-8u171-linux-x64.rpm"
             # ignore_errors: true
    when: x.failed== true

  - name: "Namenode copying file of hadoop"
    copy:
        src: "/hadoop-1.2.1-1.x86_64.rpm"
        dest: "/"

  - name: "Namenode Checking hadoop is present or not"
    command:
            "hadoop version"
    register: y
    ignore_errors: true

  - name: "Namenode installing hadoop"
    command:
            "rpm -i /hadoop-1.2.1-1.x86_64.rpm --force"
    when: x.failed== true
```

2. Creating directory for master node and updating the /etc/hadoop/hdfs-site.xml and /etc/hadoop/core-site.xml file

```
  - name: "create file for masternode"
    file:
            name: "{{dir}}"
            state: directory
    register: z

  - name: "Namenode configuration of hdfs-site.xml"
    blockinfile:
      path: "/etc/hadoop/hdfs-site.xml"
      insertafter: '<configuration>'
      block: "<property>\n<name>dfs.name.dir</name>\n<value>{{dir}}</value>\n</proper
ty>"


  - name: "Namenode configuration of core-site.xml"
    blockinfile:
      path: "/etc/hadoop/core-site.xml"
      insertafter: '<configuration>'
      block: "<property>\n<name>fs.default.name</name>\n<value>hdfs://{{ansible_facts
['default_ipv4']['address']}}:{{por}}</value>\n</property>"
```

3. Format the /master directory to store metadata of data nodes. then start service of master node.

```
- name: "Namenode formating directory"
  command:
          "echo Y | hadoop namenode -format"
  when: z.failed== true
#- name: "Namenode checking hadoop service is running or not"

- name: "Namenode start hadoop service"
  command:
          "hadoop-daemon.sh start namenode"
```

4. Then firewall rules like 50070/tcp,50010/tcp and 9001/tcp because 9001 is used for service and 50070 is used for WebUI.

```
- name: "Namenode rules"
  firewalld:
         port: 50070/tcp
         state: enabled
         immediate: yes


- name: "Namenode rules"
  firewalld:
         port: 50010/tcp
         state: enabled
         immediate: yes
- name: "Namenode rules"
  firewalld:
         port: "{{por}}/tcp"
         state: enabled
         immediate: yes
- name: "namenode services"
  command: "jps"
  register: disp
- debug:
       var: disp
```

lets start, configuration of datanode

1.Transfer JDK and install it on target node because Hadoop built from java language then transfer Hadoop library which will be compactiable with java.

```
- hosts: slave
  vars_files:
        - vars.yml
  tasks:
  - name: "Datanode copying file of hotspot java"
    copy:
        src: "/jdk-8u171-linux-x64.rpm"
        dest: "/"


  - name: "Datanode checking java is present or not"
    command:
          "jps"
    register: x
    ignore_errors: true

  - name: "Datanode installing java"
    command:
          "rpm -i /jdk-8u171-linux-x64.rpm"
          # ignore_errors: true
    when: x.failed== true
```

```
- name: "Datanode copying file of hadoop"
  copy:
      src: "/hadoop-1.2.1-1.x86_64.rpm"
      dest: "/"

- name: "Datanode Checking hadoop is present or not"
  command:
          "hadoop version"
  register: y
  ignore_errors: true

- name: "Datanode installing hadoop"
  command:
         "rpm -i /hadoop-1.2.1-1.x86_64.rpm --force"
  when: x.failed== true
```

2. Creating directory for data node and updating the /etc/hadoop/hdfs-site.xml and /etc/hadoop/core-site.xml file

```
- name: "Datanode create file for datanode"
  file:
          name: "{{dir_data}}"
          state: directory
  register: z

- name: "Datanode configuration of hdfs-site.xml"
  blockinfile:
    path: "/etc/hadoop/hdfs-site.xml"
    insertafter: '<configuration>'
    block: "<property>\n<name>dfs.data.dir</name>\n<value>{{dir_data}}</value>\n</p
operty>"


- name: "Datanode configuration of core-site.xml"
  blockinfile:
    path: "/etc/hadoop/core-site.xml"
    insertafter: '<configuration>'
    block: "<property>\n<name>fs.default.name</name>\n<value>hdfs://{{IP}}:{{por}}<
/value>\n</property>"
```

3.start service of data node and add the firewall rules

```
- name: "Datanode checking hadoop service is running or not"
  shell:
        ps -ef | grep hadoop |grep -P 'datanode'
  register: disp

- debug:
      var: disp
- name: "Datanode start hadoop service"
  command:
          "hadoop-daemon.sh start datanode"

- name: "Datanode rule"
  firewalld:
        port: 50010/tcp
        state: enabled
        immediate: yes
- name: "Datanode rules"
  firewalld:
        port: "{{por}}/tcp"
        state: enabled
        immediate: yes
```

then lets create file for variables which which was mentioned in above playbook

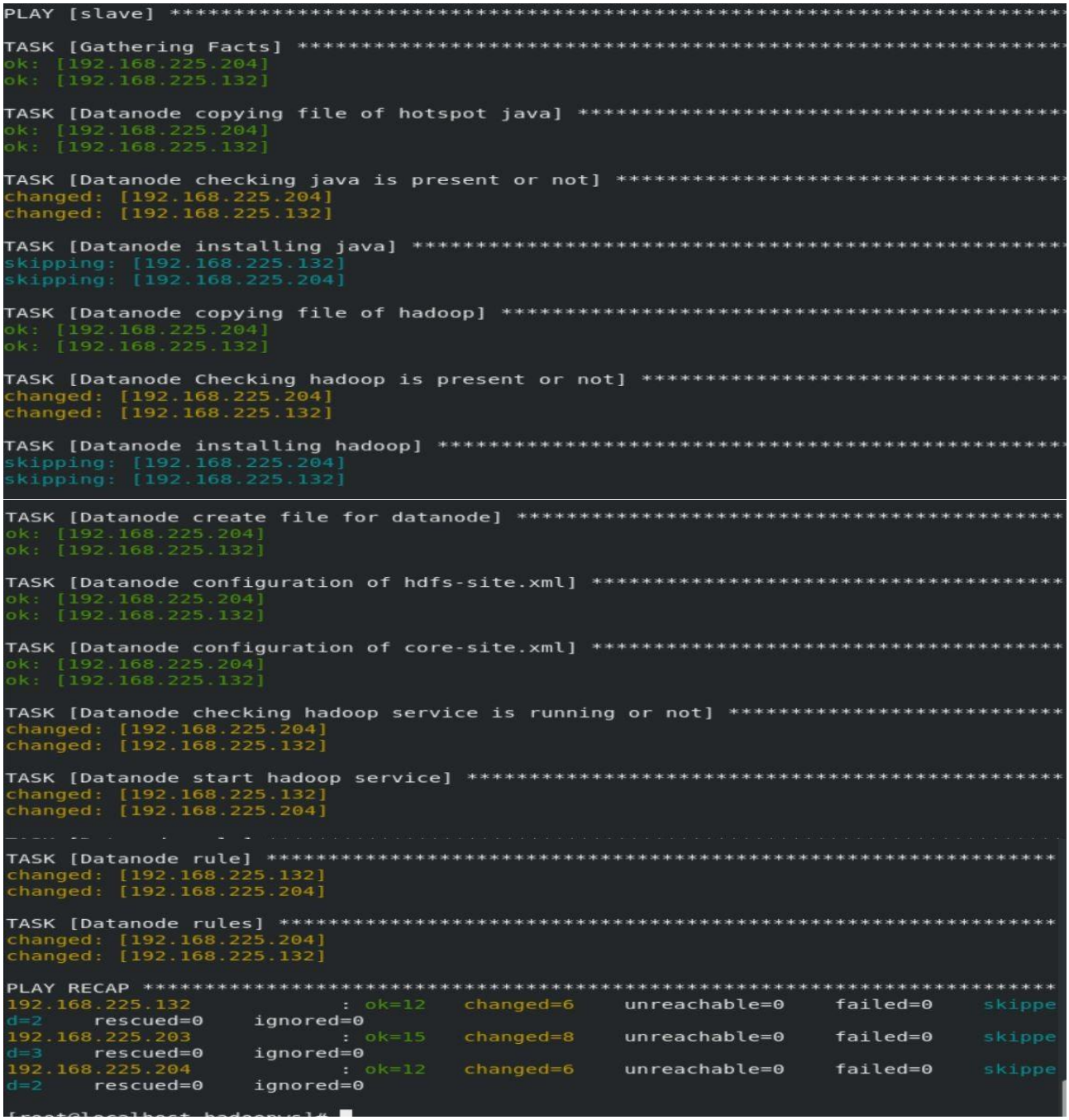
```
File  Edit  View  Search  Terminal  Help
[root@localhost hadoopws]# ls
hadoop.yml  hadoop.yml~  vars.yml
[root@localhost hadoopws]# vim vars.yml
[root@localhost hadoopws]# cat vars.yml
dir: /master
por: 9001
dir_data: /slave
IP: 192.168.225.203

[root@localhost hadoopws]#
```

then check syntax of playbook by **ansible-playbook --syntax-check playbook_name** then run playbook by ansible-playbook playbook_name

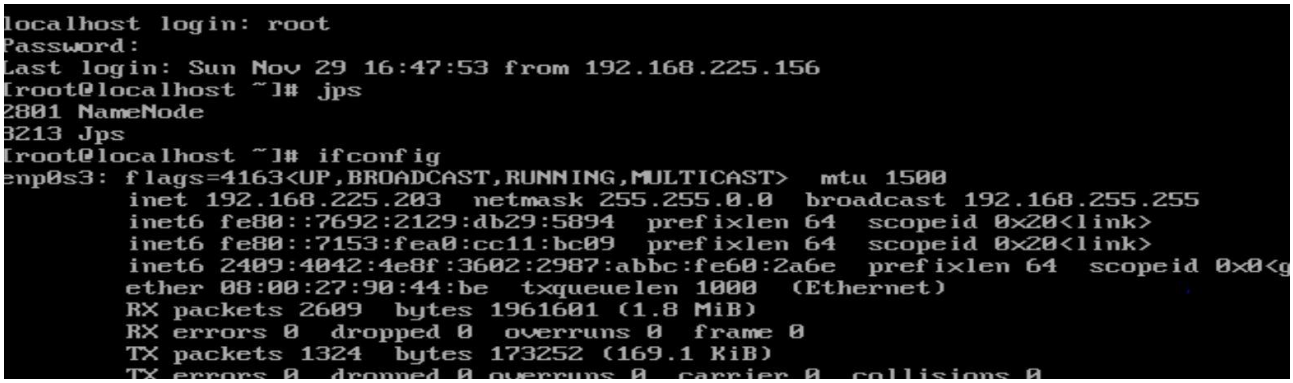
```
[root@localhost hadoopws]# ls
hadoop.yml  hadoop.yml~  vars.yml
[root@localhost hadoopws]# ansible-playbook hadoop.yml
^C [ERROR]: User interrupted execution
[root@localhost hadoopws]# ansible-playbook --syntax-check hadoop.yml
[DEPRECATION WARNING]: The firewalld module has been moved to the ansible.posix
collection. This feature will be removed from community.general in version 2.0.0.
Deprecation warnings can be disabled by setting deprecation_warnings=False in
ansible.cfg.

playbook: hadoop.yml
[root@localhost hadoopws]#
```

```
[root@localhost hadoopws]# ls
hadoop.yml  hadoop.yml~  vars.yml
[root@localhost hadoopws]# ansible all -m ping
192.168.225.132 | SUCCESS => {
    "ansible_facts": {
        "discovered_interpreter_python": "/usr/libexec/platform-python"
    },
    "changed": false,
    "ping": "pong"
}
192.168.225.204 | SUCCESS => {
    "ansible_facts": {
        "discovered_interpreter_python": "/usr/libexec/platform-python"
    },
    "changed": false,
    "ping": "pong"
}
192.168.225.203 | SUCCESS => {
    "ansible_facts": {
        "discovered_interpreter_python": "/usr/libexec/platform-python"
    },
    "changed": false,
    "ping": "pong"
}
[root@localhost hadoopws]#
```

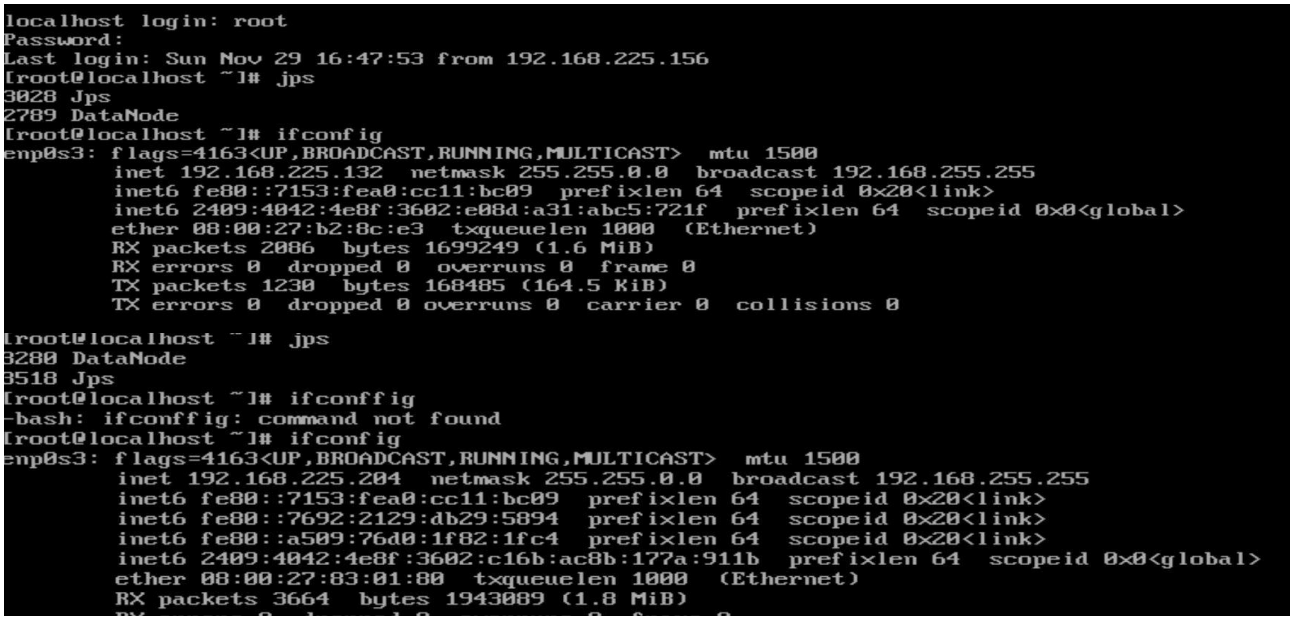
```
[root@localhost hadoopws]# ansible-playbook hadoop.yml
[DEPRECATION WARNING]: The firewalld module has been moved to the ansible.posix
collection. This feature will be removed from community.general in version 2.0.0.
Deprecation warnings can be disabled by setting deprecation_warnings=False in
ansible.cfg.

PLAY [master] ******************************************************************

TASK [Gathering Facts] *********************************************************
ok: [192.168.225.203]

TASK [Namenode copying file of hotspot java] ***********************************
ok: [192.168.225.203]

TASK [Namenode checking java is present or not] ********************************
changed: [192.168.225.203]

TASK [Namenode installing java] ************************************************
skipping: [192.168.225.203]

TASK [Namenode copying file of hadoop] *****************************************
ok: [192.168.225.203]

TASK [Namenode Checking hadoop is present or not] ******************************
changed: [192.168.225.203]

TASK [Namenode installing hadoop] **********************************************
skipping: [192.168.225.203]

TASK [create file for masternode] **********************************************
ok: [192.168.225.203]

TASK [Namenode configuration of hdfs-site.xml] *********************************
ok: [192.168.225.203]

TASK [Namenode configuration of core-site.xml] *********************************
ok: [192.168.225.203]

TASK [Namenode formating directory] ********************************************
skipping: [192.168.225.203]

TASK [Namenode start hadoop service] *******************************************
changed: [192.168.225.203]

TASK [Namenode start hadoop service] *******************************************
changed: [192.168.225.203]

TASK [Namenode rules] **********************************************************
changed: [192.168.225.203]

TASK [Namenode rules] **********************************************************
changed: [192.168.225.203]

TASK [Namenode rules] **********************************************************
changed: [192.168.225.203]

TASK [namenode services] *******************************************************
changed: [192.168.225.203]

TASK [debug] *******************************************************************
ok: [192.168.225.203] => {
    "disp": {
        "changed": true,
        "cmd": [
            "jps"
        ],
        "delta": "0:00:00.336108",
        "end": "2020-11-29 16:49:22.001819",
        "failed": false,
        "rc": 0,
        "start": "2020-11-29 16:49:21.665711",
        "stderr": "",
        "stderr_lines": [],
        "stdout": "2801 NameNode\n3191 Jps",
        "stdout_lines": [
            "2801 NameNode",
            "3191 Jps"
        ]
    }
}
```


PLAY [slave] **
TASK [Gathering Facts] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode copying file of hotspot java] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode checking java is present or not] **
changed: [192.168.225.204]
changed: [192.168.225.132]
TASK [Datanode installing java] **
skipping: [192.168.225.132]
skipping: [192.168.225.204]
TASK [Datanode copying file of hadoop] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode Checking hadoop is present or not] **
changed: [192.168.225.204]
changed: [192.168.225.132]
TASK [Datanode installing hadoop] **
skipping: [192.168.225.204]
skipping: [192.168.225.132]
TASK [Datanode create file for datanode] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode configuration of hdfs-site.xml] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode configuration of core-site.xml] **
ok: [192.168.225.204]
ok: [192.168.225.132]
TASK [Datanode checking hadoop service is running or not] **
changed: [192.168.225.204]
changed: [192.168.225.132]
TASK [Datanode start hadoop service] **
changed: [192.168.225.132]
changed: [192.168.225.204]
TASK [Datanode rule] **
changed: [192.168.225.132]
changed: [192.168.225.204]
TASK [Datanode rules] **
changed: [192.168.225.204]
changed: [192.168.225.132]
PLAY RECAP **
192.168.225.132 : ok=12 changed=6 unreachable=0 failed=0 skippe
d=2 rescued=0 ignored=0
192.168.225.203 : ok=15 changed=8 unreachable=0 failed=0 skippe
d=3 rescued=0 ignored=0
192.168.225.204 : ok=12 changed=6 unreachable=0 failed=0 skippe
d=2 rescued=0 ignored=0

Name node:



Data Node:



Thus I have successfully completed task 4