



12/22/2020

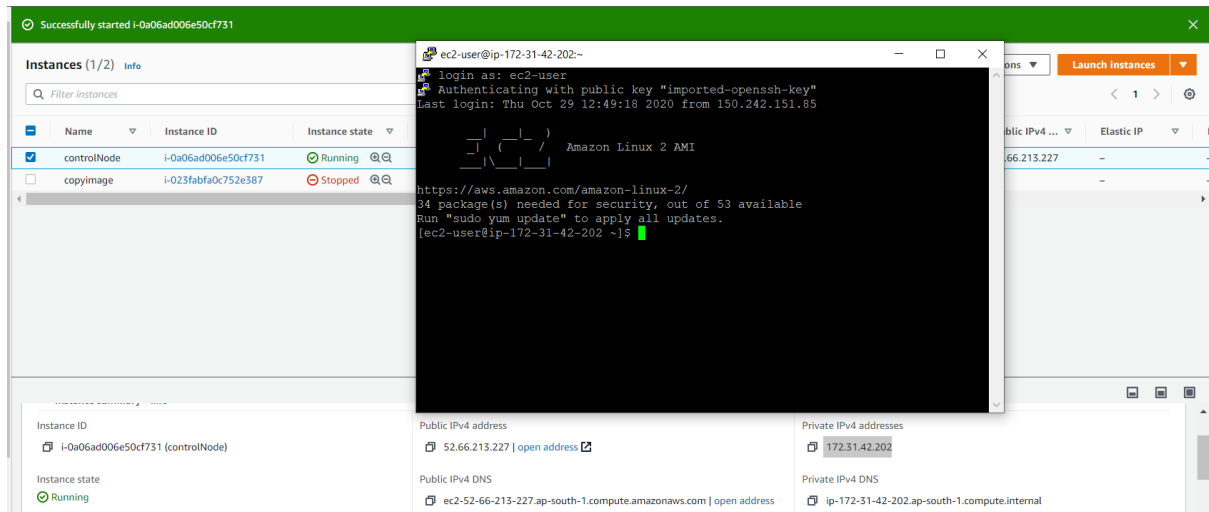
RH294 Ansible Automation, configuring Hadoop Cluster.



[akshay anil](#)

AKSHAYANIL1080.GITHUB.IO/MYWEBSITE

Login into ec2-instance via putty



1. Install ansible

#sudo amazon-linux-extras install ansible2

```
[root@ip-172-31-42-202 ~]# sudo amazon-linux-extras install ansible2
Installing ansible
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
Cleaning repos: amzn2-core amzn2extra-ansible2 amzn2extra-docker
12 metadata files removed
4 sqlite files removed
0 metadata files removed
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
amzn2-core
amzn2extra-ansible2
amzn2extra-docker
(1/7): amzn2-core/2/x86_64/group_gz
(2/7): amzn2-core/2/x86_64/updateinfo
(3/7): amzn2extra-docker/2/x86_64/primary_db
(4/7): amzn2extra-ansible2/2/x86_64/updateinfo
(5/7): amzn2extra-ansible2/2/x86_64/primary_db
```

Checking the version of ansible:

#ansible --version

```
[root@ip-172-31-42-202 ~]# ansible --version
ansible 2.9.13
  config file = /etc/ansible/ansible.cfg
  configured module search path = [u'/root/.ansible/plugins/modules', u'/usr/share/ansible/plugins/modules']
  ansible python module location = /usr/lib/python2.7/site-packages/ansible
  executable location = /usr/bin/ansible
  python version = 2.7.18 (default, Aug 27 2020, 21:22:52) [GCC 7.3.1 20180712 (Red Hat 7.3.1-9)]
```

MY INSTANCES:

MY CONTROL NODE IP: 13.127.127.215

Filter instances

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status
<input checked="" type="checkbox"/>	controlNode	i-0a06ad006e50cf731	Running	t2.micro	2/2 checks ...	No alarms
<input type="checkbox"/>	Datanode	i-02150b34a2a061415	Running	t2.micro	2/2 checks ...	No alarms
<input type="checkbox"/>	Namenode	i-0eb82dabd4caf047c	Running	t2.micro	2/2 checks ...	No alarms

Instance: i-0a06ad006e50cf731 (controlNode)

Details

Security

Networking

Storage

Status Checks

Monitoring

Tags

Networking details

Info

Public IPv4 address

13.127.127.215 | open address

Public IPv4 DNS

Private IPv4 addresses

172.31.42.202

Private IPv4 DNS

NAME NODE IP: 15.207.85.47

Filter instances

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
<input type="checkbox"/>	controlNode	i-0a06ad006e50cf731	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a
<input type="checkbox"/>	Datanode	i-02150b34a2a061415	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a
<input checked="" type="checkbox"/>	Namenode	i-0eb82dabd4caf047c	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a

Instance: i-0eb82dabd4caf047c (Namenode)

Details

Security

Networking

Storage

Status Checks

Monitoring

Tags

Networking details

Info

Public IPv4 address

15.207.85.47 | open address

Public IPv4 DNS

Private IPv4 addresses

172.31.44.57

Private IPv4 DNS

DATA NODE IP: 15.207.85.25

Filter instances

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
<input type="checkbox"/>	controlNode	i-0a06ad006e50cf731	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a
<input checked="" type="checkbox"/>	Datanode	i-02150b34a2a061415	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a
<input type="checkbox"/>	Namenode	i-0eb82dabd4caf047c	Running	t2.micro	2/2 checks ...	No alarms +	ap-south-1a

Instance: i-02150b34a2a061415 (Datanode)

Details

Security

Networking

Storage

Status Checks

Monitoring

Tags

Networking details

Info

Public IPv4 address

15.207.85.25 | open address

Private IPv4 addresses

172.31.32.152

2. ENABLING THE SSH BETWEEN CONTROL AND TARGET NODE.

Open control node ,type the following commands.

Cd .ssh

Ssh-keygen

Ls

```
[root@ip-172-31-42-202 ~]# cd .ssh
[root@ip-172-31-42-202 .ssh]# ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
/root/.ssh/id_rsa already exists.
Overwrite (y/n)?
[root@ip-172-31-42-202 .ssh]#
[root@ip-172-31-42-202 .ssh]# ls
authorized_keys  ec2-key.pem  id_rsa  id_rsa.pub  known_hosts
[root@ip-172-31-42-202 .ssh]#
```

Do the same with all taerget nodes.

Open the id_rsa.pub file from control node and copy the content into the authrization_keys of the target node .

This will enablet ssh between control and target/managed node.

3. WRITING THE CODE.

```
- hosts: all
  tasks:

    - name: "sending jdk software to targetNode"
      copy:
        dest: "/root"
        src: "/home/ec2-user/jdk-8u171-linux-x64.rpm"

    - name: "sending hadoop software to targetNode"
      copy:
        dest: "/root"
        src: "/home/ec2-user/hadoop-1.2.1-1.x86_64.rpm"

    - name: "installing Jdk software"
      command: "rpm -i jdk-8u171-linux-x64.rpm"
      ignore_errors: yes

    - name: "installing hadoop software"
      command: "rpm -i hadoop-1.2.1-1.x86_64.rpm --force"
      ignore_errors: yes

    #NAME NODE SETUP

- hosts: mynamenode

  tasks:
    - name: "deleting the previous directory"
      shell: "rm -rf /nn"
      ignore_errors: yes
    - name: "create a directory"
```

```

- name: "create a directory"
  file:
    state: directory
    path: "/nn"

- name: "hdfs-site file in namenode"
  template:
    dest: "/etc/hadoop/hdfs-site.xml"
    src: "nn_hdfs.xml"

- name: "core-site file in namenode"
  template:
    dest: "/etc/hadoop/core-site.xml"
    src: "nn_core.xml"
- name: "formatting the namenode"
  shell: "echo Y | hadoop namenode -format"


- name: "stopping namenode service"
  command: "hadoop-daemon.sh stop namenode"

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


  #DATA NODE
- hosts: mydatanode

  tasks:

- name: "deleting the previous directory"
  shell: "rm -rf /dn"
  ignore_errors: yes

- name: "create a directory"
  file:
    state: directory

```

```
    path: "/dn"

- name: "hdfs-site file in datanode"
  template:
    dest: "/etc/hadoop/hdfs-site.xml"
    src: "dn_hdfs.xml"

- name: "core-site file in datanode"
  template:
    dest: "/etc/hadoop/core-site.xml"
    src: "dn_core.xml"
- name: "stopping datanode service"
  command: "hadoop-daemon.sh stop datanode"
  ignore_errors: yes

- name: "starting datanode service"
  command: "hadoop-daemon.sh start datanode"
```

Running

```
[root@ip-172-31-42-202 ws]# ansible-playbook preq.yml

PLAY [all] *****

TASK [Gathering Facts] *****
[WARNING]: Platform linux on host 15.207.85.47 is using the discovered Python interpreter at /usr/bin/python, but future installation of another Python interpreter could change this. See https://docs.ansible.com/ansible/2.9/referen
ce_appendices/interpreter_discovery.html for more information.
ok: [15.207.85.47]
[WARNING]: Platform linux on host 15.207.85.25 is using the discovered Python interpreter at /usr/bin/python, but future installation of another Python interpreter could change this. See https://docs.ansible.com/ansible/2.9/referen
ce_appendices/interpreter_discovery.html for more information.
ok: [15.207.85.25]

TASK [sending jdk software to targetNode] *****
ok: [15.207.85.25]
ok: [15.207.85.47]

TASK [sending hadoop software to targetNode] *****
ok: [15.207.85.25]
ok: [15.207.85.47]

TASK [installing hadoop software] *****
changed: [15.207.85.47]
changed: [15.207.85.25]

PLAY [mynamenode] *****

TASK [Gathering Facts] *****
ok: [15.207.85.47]

TASK [deleting the previous directory] *****
[WARNING]: Consider using the file module with state=absent rather than running
'rm'. If you need to use command because file is insufficient you can add
'warn: false' to this command task or set 'command_warnings=False' in
ansible.cfg to get rid of this message.
changed: [15.207.85.47]

TASK [create a directory] *****
changed: [15.207.85.47]

TASK [hdfs-site file in namenode] *****
```



```

TASK [core-site file in namenode] *****
*****
ok: [15.207.85.47]

TASK [formatting the namemode] *****
*****
changed: [15.207.85.47]

TASK [stopping namenode service] *****
*****
changed: [15.207.85.47]

TASK [starting namenode service] *****
*****
changed: [15.207.85.47]

PLAY [mydatanode] *****
*****

TASK [Gathering Facts] *****
*****
ok: [15.207.85.25]

TASK [deleting the previous directory] *****
*****
changed: [15.207.85.25]

TASK [craete a directory] *****
*****
changed: [15.207.85.25]

TASK [hdfs-site file in datanode] *****
*****

```

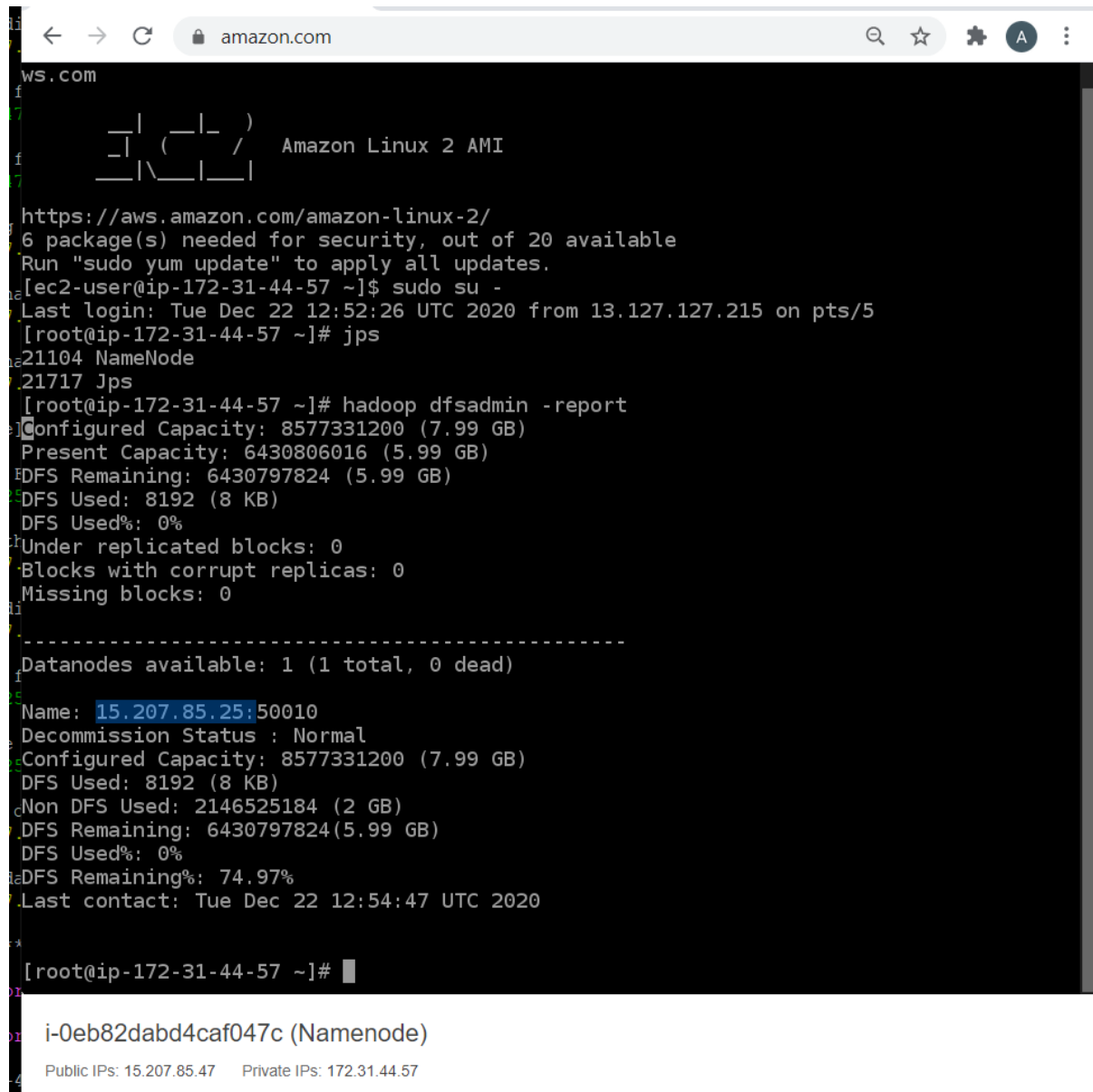
```

PLAY RECAP *****
*****
15.207.85.25      : ok=12    changed=6    unreachable=0
  failed=0      skipped=0    rescued=0    ignored=1
15.207.85.47     : ok=13    changed=7    unreachable=0
  failed=0      skipped=0    rescued=0    ignored=1

```

VERIFICATION

Name node status



```
aws.com
f
7
_ | ( _ | _ )
_ | / Amazon Linux 2 AMI
_ | \ _ | _ |

https://aws.amazon.com/amazon-linux-2/
6 package(s) needed for security, out of 20 available
Run "sudo yum update" to apply all updates.
[ec2-user@ip-172-31-44-57 ~]$ sudo su -
Last login: Tue Dec 22 12:52:26 UTC 2020 from 13.127.127.215 on pts/5
[root@ip-172-31-44-57 ~]# jps
21104 NameNode
21717 Jps
[root@ip-172-31-44-57 ~]# hadoop dfsadmin -report
Configured Capacity: 8577331200 (7.99 GB)
Present Capacity: 6430806016 (5.99 GB)
DFS Remaining: 6430797824 (5.99 GB)
DFS Used: 8192 (8 KB)
DFS Used%: 0%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

-----
Datanodes available: 1 (1 total, 0 dead)
Name: 15.207.85.25:50010
Decommission Status : Normal
Configured Capacity: 8577331200 (7.99 GB)
DFS Used: 8192 (8 KB)
Non DFS Used: 2146525184 (2 GB)
DFS Remaining: 6430797824 (5.99 GB)
DFS Used%: 0%
DFS Remaining%: 74.97%
Last contact: Tue Dec 22 12:54:47 UTC 2020

[root@ip-172-31-44-57 ~]# █

i-0eb82dabd4caf047c (Namenode)
Public IPs: 15.207.85.47 Private IPs: 172.31.44.57
```

Data Node status

```
Last login: Tue Dec 22 12:42:30 2020 from ec2-13-233-177-1.ap-south
aws.com

  _ | _ | _ )
  _ | ( _ | /   Amazon Linux 2 AMI
  __ |\__ |__ |

https://aws.amazon.com/amazon-linux-2/
6 package(s) needed for security, out of 20 available
Run "sudo yum update" to apply all updates.
[ec2-user@ip-172-31-32-152 ~]$ sudo su -
Last login: Tue Dec 22 12:52:54 UTC 2020 from 13.127.127.215 on pts
[root@ip-172-31-32-152 ~]# jps
19298 Jps
18727 DataNode
[root@ip-172-31-32-152 ~]# █
```

i-02150b34a2a061415 (Datanode)

Public IPs: 15.207.85.25 Private IPs: 172.31.32.152

AWS GUI Checking

← → ↻ ⚠ Not secure | 15.207.85.47:50070/dfshealth.jsp

NameNode '0.0.0.0:9001'

Started: Tue Dec 22 12:52:27 UTC 2020
Version: 1.2.1, r1503152
Compiled: Mon Jul 22 15:27:42 PDT 2013 by mattf
Upgrades: There are no upgrades in progress.

[Browse the filesystem](#)
[Namenode Logs](#)

Cluster Summary

1 files and directories, 0 blocks = 1 total. Heap Size is 15.5 MB / 123.75 MB (12%)

Configured Capacity : 7.99 GB
DFS Used : 8 KB
Non DFS Used : 2 GB
DFS Remaining : 5.99 GB
DFS Used% : 0 %
DFS Remaining% : 74.97 %
[Live Nodes](#) : 1
[Dead Nodes](#) : 0
[Decommissioning Nodes](#) : 0
Number of Under-Replicated Blocks : 0

NameNode Storage:

Storage Directory	Type	State
/nn	IMAGE_AND_EDITS	Active

This is [Apache Hadoop](#) release 1.2.1

ALL RUNNING WELL.....