

Commissioning and Decommissioning data nodes

A guide to Add the data nodes or remove the data nodes from the cluster

edureka!

edureka!

Software Requirements

- ✓ VMware Player or Oracle Virtual Box
- ✓ CentOS Virtual Machine

Hardware Requirements

- ✓ Intel Core i3 processor or higher
- ✓ **8 GB RAM Recommended**
- ✓ **300 GB for VM Recommended (By default 40 GB is taken)**

edureka!

Introduction

This setup and configuration document is a guide to install Commissioning and Decommissioning data nodes from the Multi-Node Apache Hadoop 1.2.1 cluster on a CentOS virtual machine on your PC.

The guide describes the whole process in two parts:

[Section 1: Commissioning Data nodes](#)

In this section we describe how to add the new data nodes to the cluster, what are the properties should be add for the commissioning data nodes.

[Section 2: De-commissioning Data nodes](#)

In this section we describe how to add the remove the data nodes from the cluster, what are the properties should be add for the de-commissioning data nodes.

Note: The configuration described here is intended for learning purposes only.

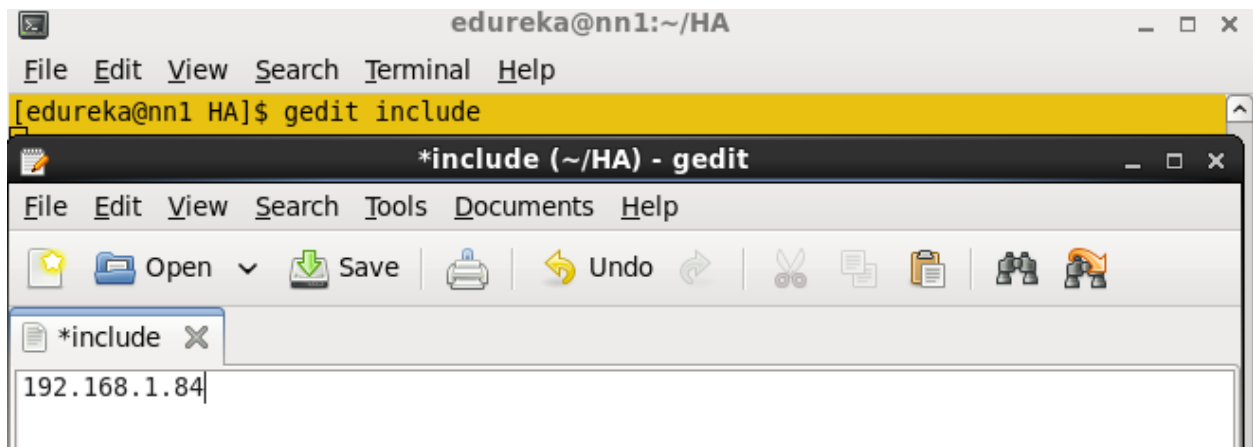
Important: Whenever you change any configuration xml files in the hadoop cluster, you have to stop the daemons and restart the daemons to apply the changes to the hadoop cluster, so better to add all the Properties (**mapred.hosts.exclude**, **mapred.hosts.include**, **hosts.include**, **hosts.exclude**) before starting the daemons.

Step-1: Commissioning Data nodes

1.1: Create file in a name node to store the IP address or Hostnames of the commissioning data nodes.

Create the file with the name include.

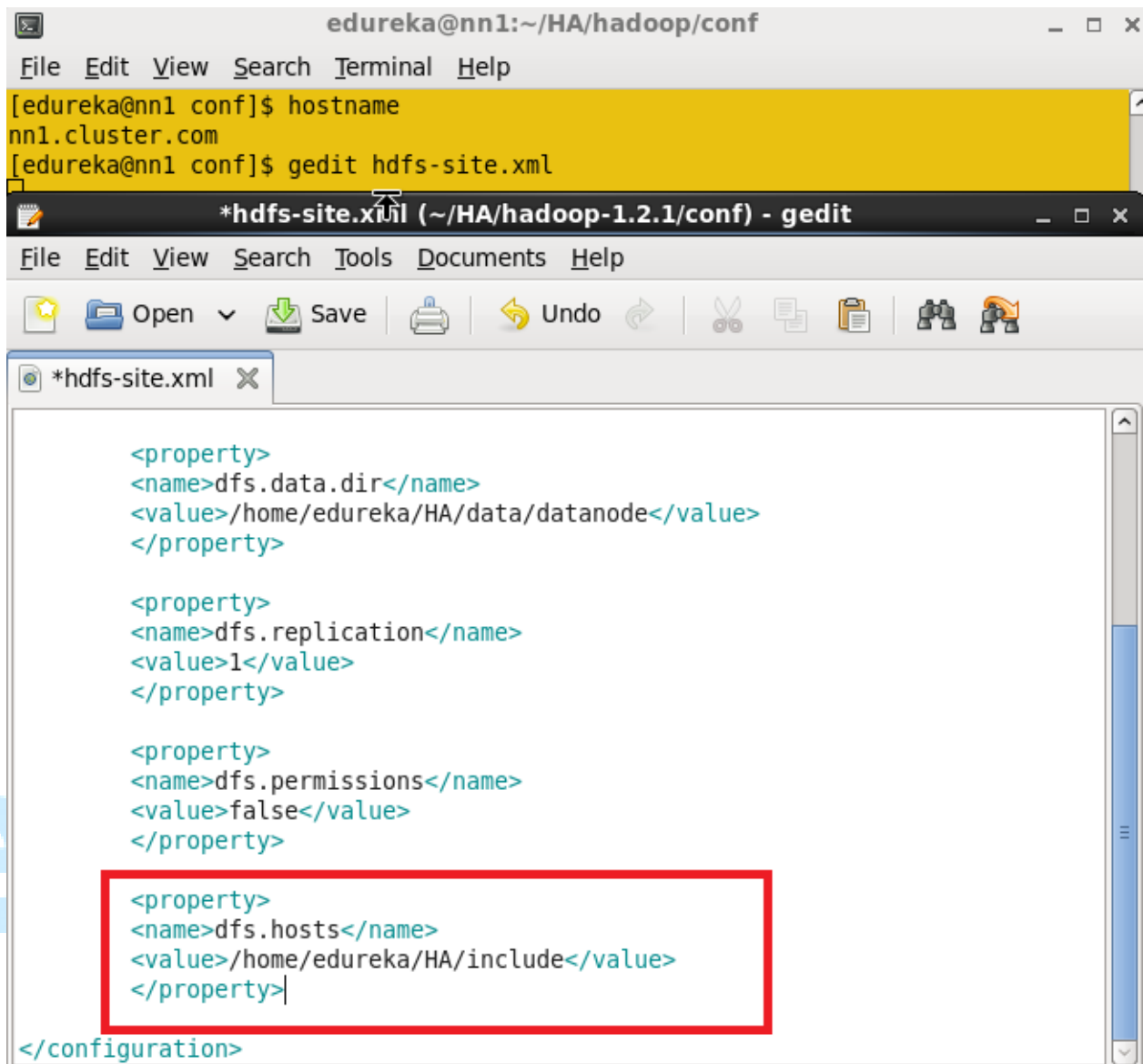
Open the include file and add the IP address of the New Data node.



1.2: Open the Name node's hdfs-site.xml file.

Add the below properties.

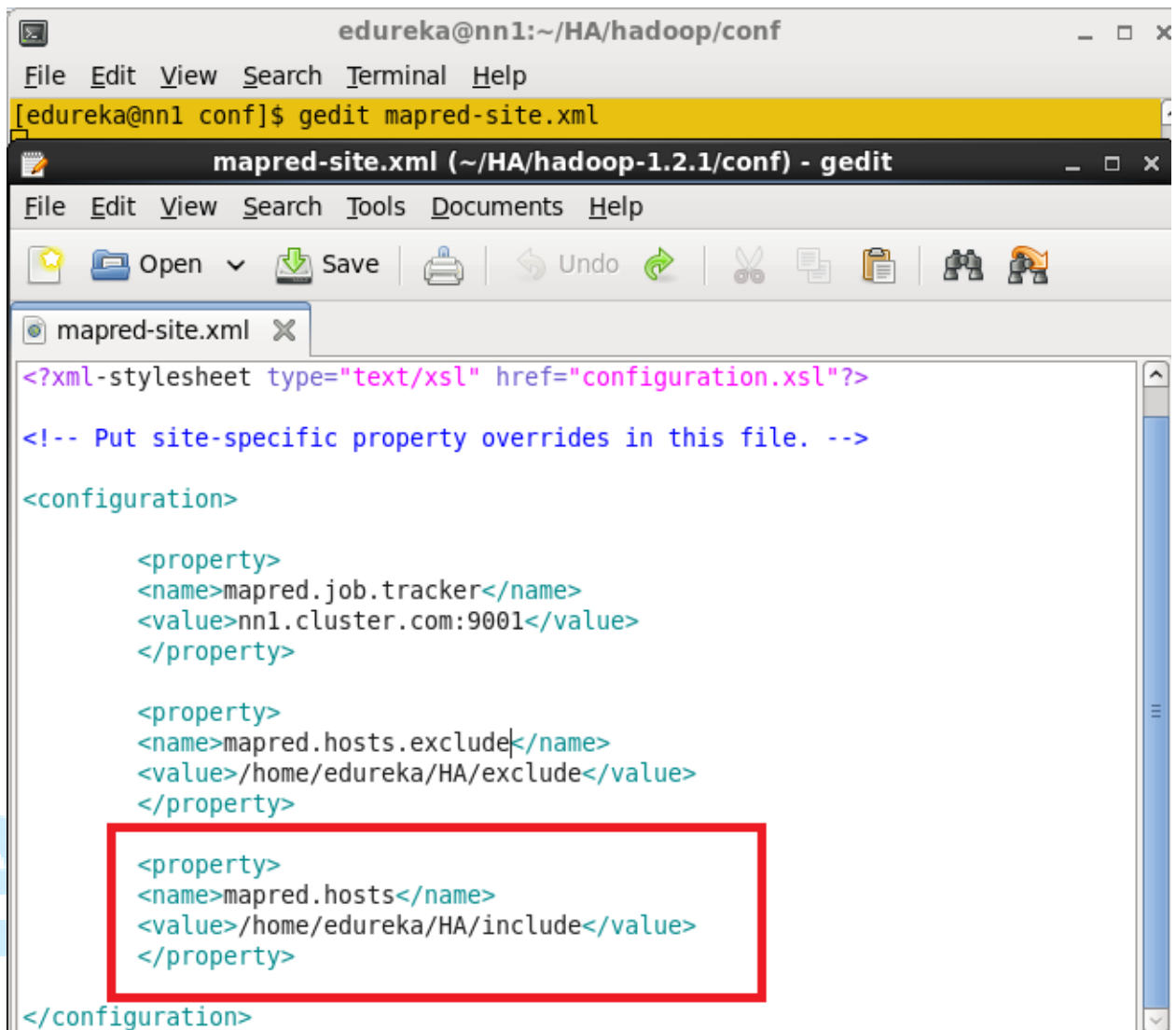
```
<property>
  <name>dfs.hosts</name>
  <value> <Path to the include file> </value>
  <final>true</final>
</property>
```



1.3: Open the mapred-site.xml file in name node.

Add the below properties.

```
<property>
  <name>mapred.hosts</name>
  <value> <Path to the include file> </value>
  <final>true</final>
</property>
```



```
edureka@nn1:~/HA/hadoop/conf
File Edit View Search Terminal Help
[edureka@nn1 conf]$ gedit mapred-site.xml

mapred-site.xml (~/HA/hadoop-1.2.1/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
mapred-site.xml x
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>nn1.cluster.com:9001</value>
  </property>
  <property>
    <name>mapred.hosts.exclude</name>
    <value>/home/edureka/HA/exclude</value>
  </property>
  <property>
    <name>mapred.hosts</name>
    <value>/home/edureka/HA/include</value>
  </property>
</configuration>
```

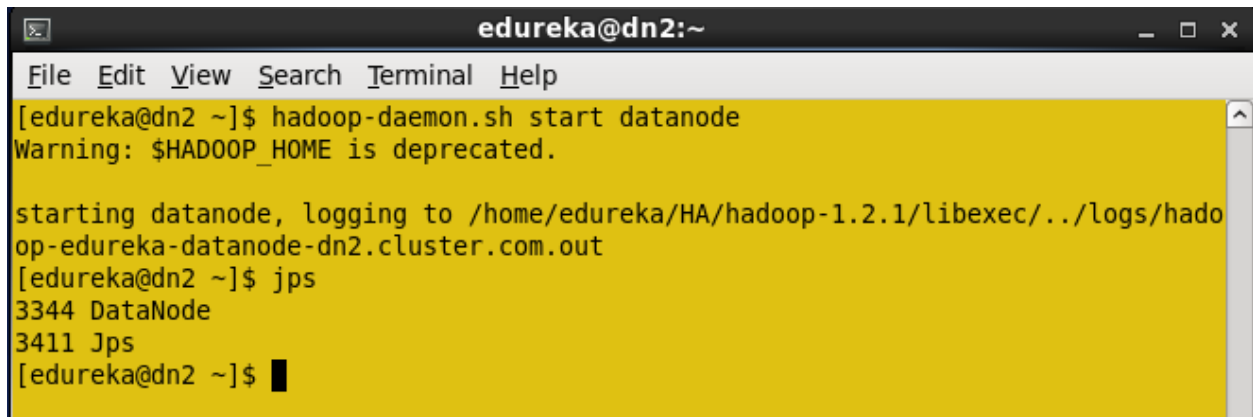
1.4: In a cluster if Name node is running in master machine stop the name node or start it, or start the cluster (start all the daemons).

Command to stop: `hadoop-daemon.sh stop namenode`

Command to start: `hadoop-daemon.sh start namenode`

1.5: Start the Datanode daemon in new datanode virtual machines and enter jps command to check the daemon.

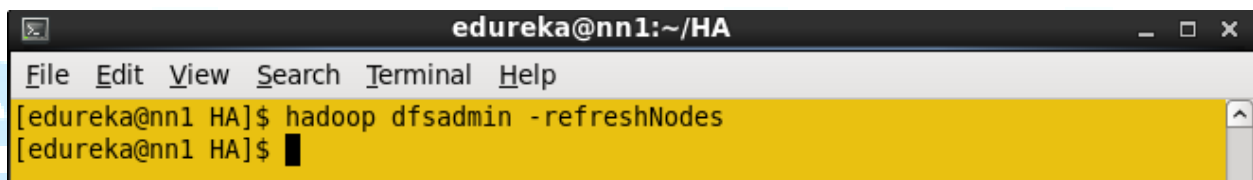
Command: `hadoop-daemon.sh start datanode`



```
edureka@dn2:~  
File Edit View Search Terminal Help  
[edureka@dn2 ~]$ hadoop-daemon.sh start datanode  
Warning: $HADOOP_HOME is deprecated.  
  
starting datanode, logging to /home/edureka/HA/hadoop-1.2.1/libexec/./logs/hadoop-edureka-datanode-dn2.cluster.com.out  
[edureka@dn2 ~]$ jps  
3344 DataNode  
3411 Jps  
[edureka@dn2 ~]$
```

1.6: Run below command in namenode virtual machine to give the information about the new datanode or allow to read the include file IP address.

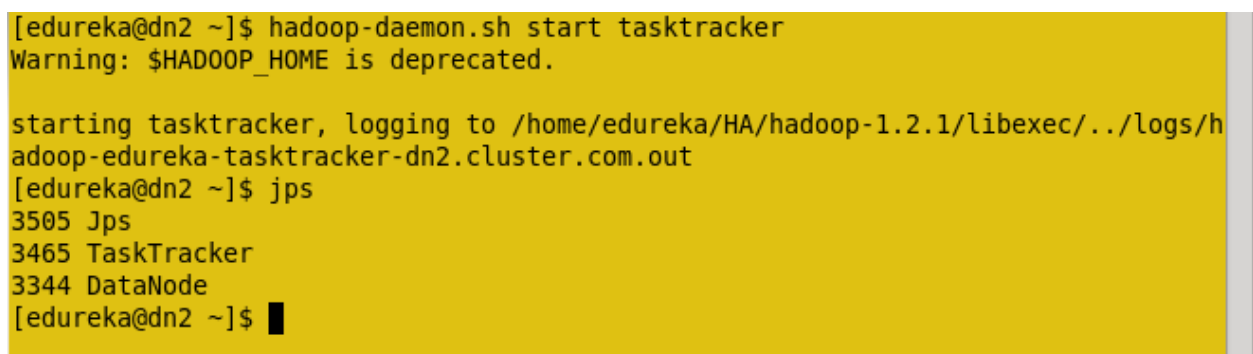
Command: `hadoop dfsadmin -refreshNodes`



```
edureka@nn1:~/HA  
File Edit View Search Terminal Help  
[edureka@nn1 HA]$ hadoop dfsadmin -refreshNodes  
[edureka@nn1 HA]$
```

1.7: Start the Tasktracker daemon in new datanode virtual machines and enter jps command to check the daemon.

Command: `hadoop-daemon.sh start tasktracker`



```
[edureka@dn2 ~]$ hadoop-daemon.sh start tasktracker  
Warning: $HADOOP_HOME is deprecated.  
  
starting tasktracker, logging to /home/edureka/HA/hadoop-1.2.1/libexec/./logs/hadoop-edureka-tasktracker-dn2.cluster.com.out  
[edureka@dn2 ~]$ jps  
3505 Jps  
3465 TaskTracker  
3344 DataNode  
[edureka@dn2 ~]$
```

1.8: Run below command in name node virtual machine to give the information about the new task tracker details.

Command: `hadoop mradmin -refreshNodes`

```
edureka@nn1:~/HA
File Edit View Search Terminal Help
[edureka@nn1 HA]$ hadoop mradmin -refreshNodes
[edureka@nn1 HA]$
```

1.9: Enter Hadoop report command in name node to check the newly added data node details.

```
edureka@nn1:~/HA
File Edit View Search Terminal Help
[edureka@nn1 HA]$ hadoop dfsadmin -report
```

```
-----
Datanodes available: 3 (3 total, 0 dead)

Name: 192.168.1.81:50010
Decommission Status : Normal
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 282624 (276 KB)
Non DFS Used: 3752357888 (3.49 GB)
DFS Remaining: 51981742080(48.41 GB)
DFS Used%: 0%
DFS Remaining%: 93.27%
Last contact: Wed Apr 29 23:10:30 CEST 2015

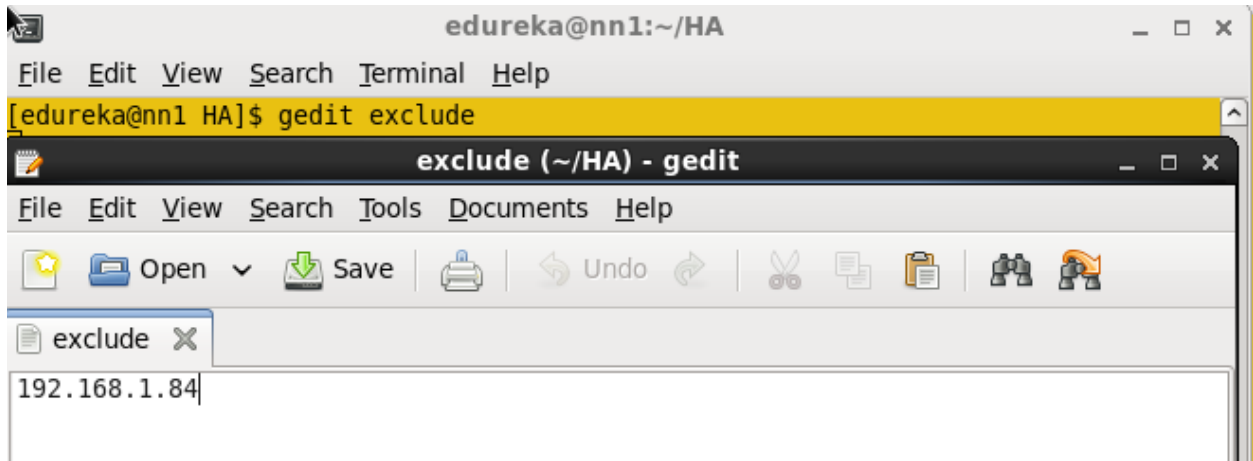
Name: 192.168.1.82:50010
Decommission Status : Normal
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 299008 (292 KB)
Non DFS Used: 3794771968 (3.53 GB)
DFS Remaining: 51939311616(48.37 GB)
DFS Used%: 0%
DFS Remaining%: 93.19%
Last contact: Wed Apr 29 23:10:30 CEST 2015

Name: 192.168.1.84:50010
Decommission Status : Normal
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 172697 (168.65 KB)
Non DFS Used: 3492392295 (3.25 GB)
DFS Remaining: 52241817600(48.65 GB)
DFS Used%: 0%
DFS Remaining%: 93.73%
-----
```


Step-2: De-commissioning Data nodes

2.1: Create file in a name node to store the IP address or Hostnames of the de-commissioning data nodes.

Create the file with the name exclude.

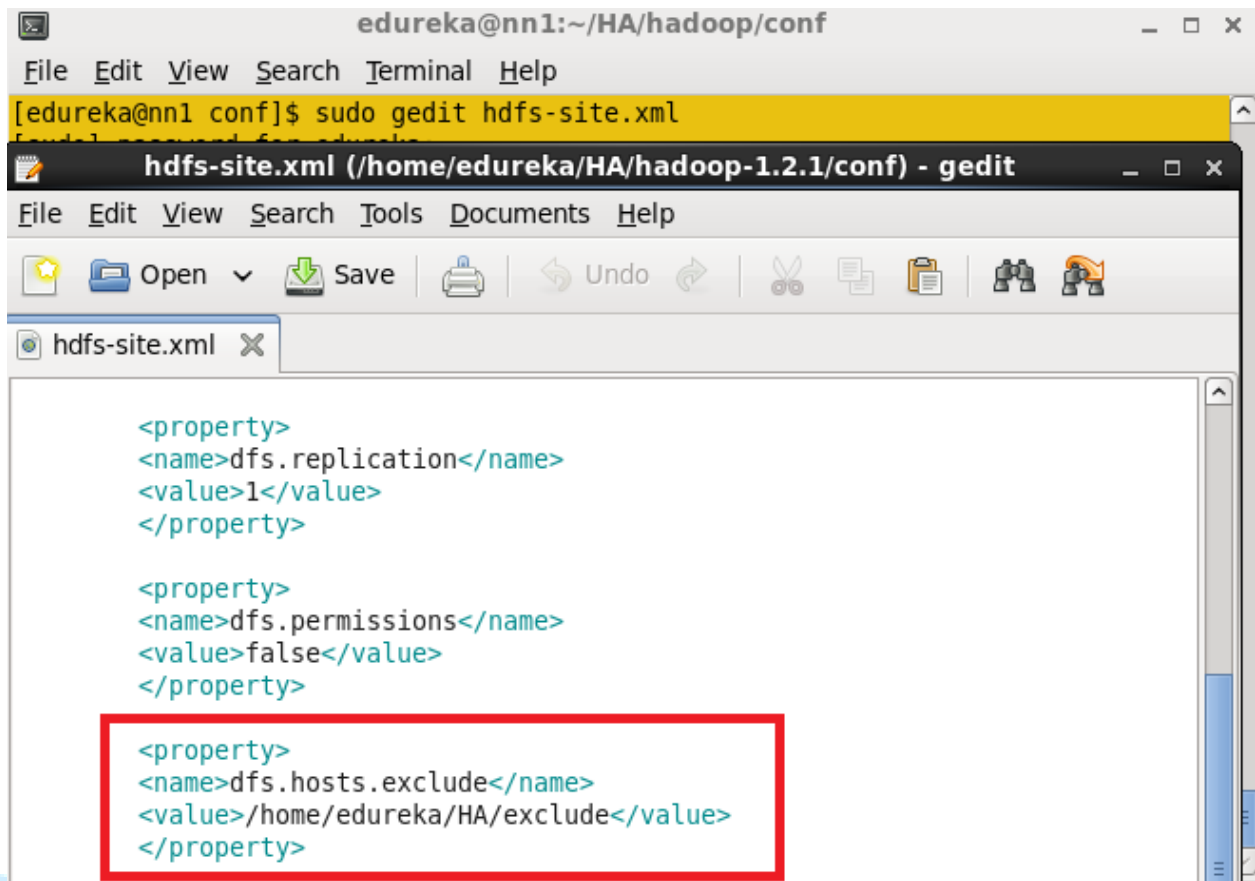


2.2: Open the exclude file and add the IP address of the Data node to de-commission.

Open the Name node's hdfs-site.xml file.

Add the below properties.

```
<property>
  <name>dfs.hosts.exclude</name>
  <value> <Path to the exclude file> </value>
  <final>true</final>
</property>
```

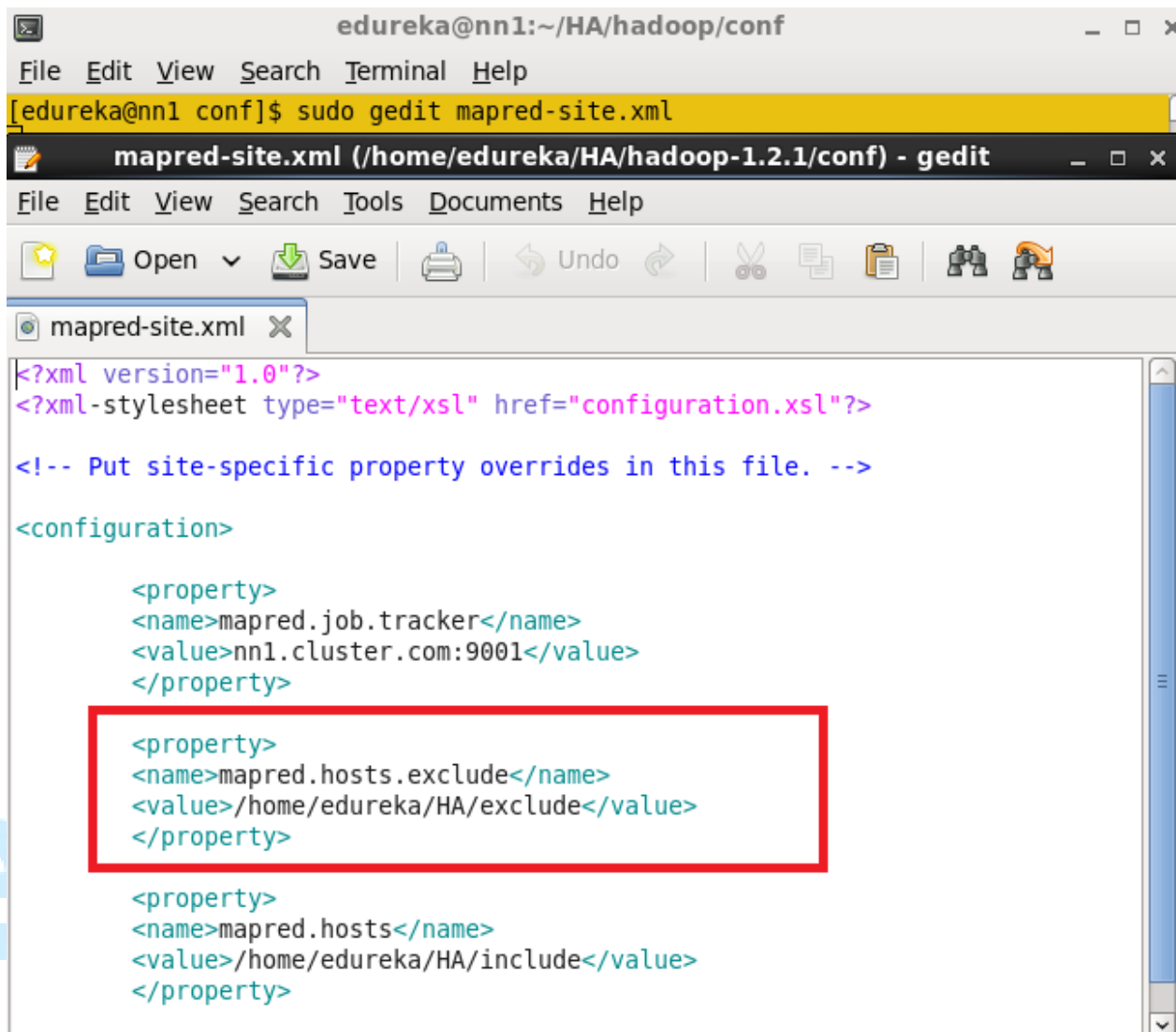


edureka!

2.3: Open the mapred-site.xml file in name node.

Add the below properties.

```
<property>
  <name>mapred.hosts.exclude</name>
  <value> <Path to the exclude file> </value>
  <final>true</final>
</property>
```



```
edureka@nn1:~/HA/hadoop/conf
File Edit View Search Terminal Help
[edureka@nn1 conf]$ sudo gedit mapred-site.xml

mapred-site.xml (/home/edureka/HA/hadoop-1.2.1/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
mapred-site.xml x
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

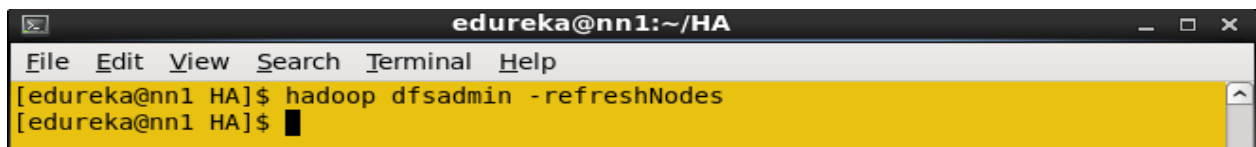
    <property>
    <name>mapred.job.tracker</name>
    <value>nn1.cluster.com:9001</value>
    </property>

    <property>
    <name>mapred.hosts.exclude</name>
    <value>/home/edureka/HA/exclude</value>
    </property>

    <property>
    <name>mapred.hosts</name>
    <value>/home/edureka/HA/include</value>
    </property>
```

2.4: Run below command in namenode virtual machine to give the information about the de-commissioning datanode or allow to read the exclude file IP address.

Command: `hadoop dfsadmin -refreshNodes`



```
edureka@nn1:~/HA
File Edit View Search Terminal Help
[edureka@nn1 HA]$ hadoop dfsadmin -refreshNodes
[edureka@nn1 HA]$
```

2.5: Enter Hadoop report command in name node to check the newly added data node details.

```
edureka@nn1:~/HA
File Edit View Search Terminal Help
[edureka@nn1 HA]$ hadoop dfsadmin -report

-----
Datanodes available: 2 (3 total, 1 dead)

Name: 192.168.1.81:50010
Decommission Status : Normal
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 282624 (276 KB)
Non DFS Used: 3752423424 (3.49 GB)
DFS Remaining: 51981676544(48.41 GB)
DFS Used%: 0%
DFS Remaining%: 93.27%
Last contact: Wed Apr 29 23:22:18 CEST 2015

Name: 192.168.1.82:50010
Decommission Status : Normal
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 299008 (292 KB)
Non DFS Used: 3794771968 (3.53 GB)
DFS Remaining: 51939311616(48.37 GB)
DFS Used%: 0%
DFS Remaining%: 93.19%
Last contact: Wed Apr 29 23:22:18 CEST 2015

Name: 192.168.1.84:50010
Decommission Status : Decommissioned
Configured Capacity: 55734382592 (51.91 GB)
DFS Used: 188416 (184 KB)
Non DFS Used: 3492388864 (3.25 GB)
DFS Remaining: 52241805312(48.65 GB)
DFS Used%: 0%
DFS Remaining%: 93.73%
```