Multi Node Hadoop Cluster Setup

This document describes how to create Hadoop Multi Node cluster in just 30 Minutes on Amazon EC2 cloud. You will learn following topics.

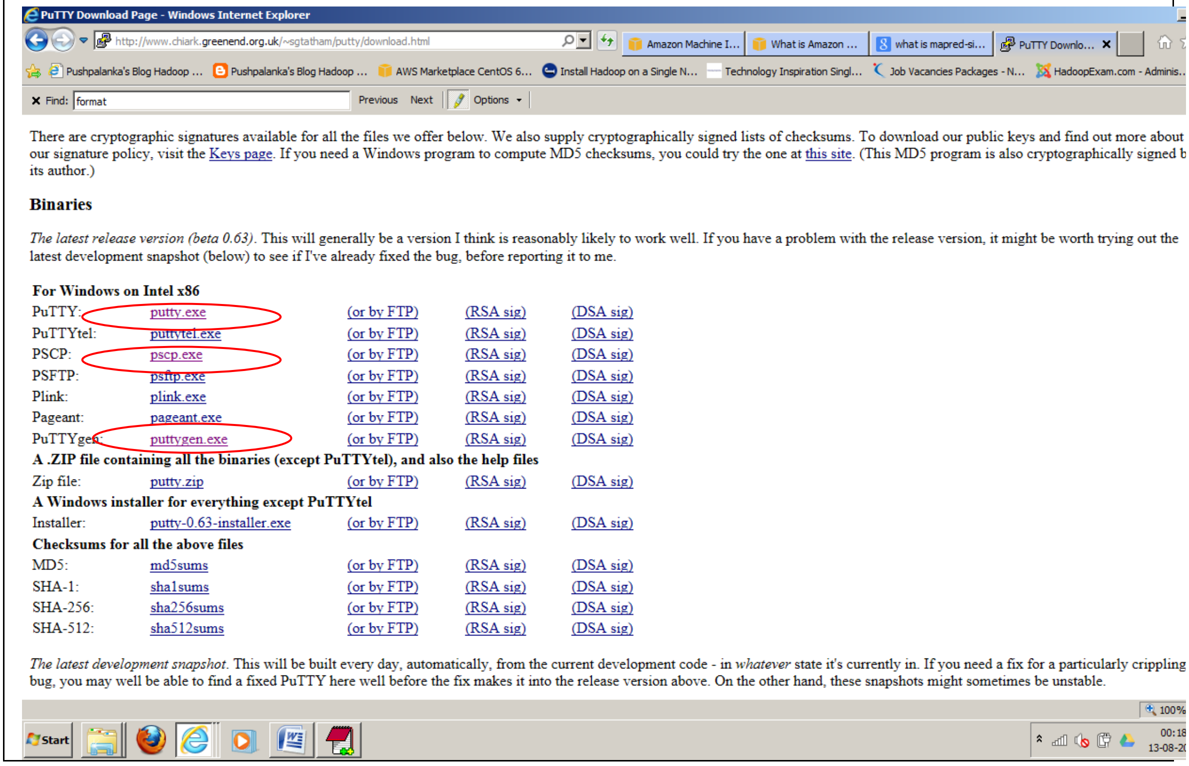
Click Here to watch these steps in Video Instructions

- How to create instance on Amazon EC2
- How to connect that Instance Using putty
- Installing Hadoop framework on this instance
- Run **NGram(End to End Project Creation already explained in Module 15)** example which come with Hadoop framework.
- Setting the replication factor

Following Software require on your local windows machine

1. Putty: To connect amazo ec2 instance.
2. puttygen: create private key from .ppm file
3. pscp : to copy file from your local filesytem to amazon instance

Download all three tools from
http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html


PuTTY Download Page - Windows Internet Explorer
http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html
Amazon Machine I...
What is Amazon ...
what is mapred-si...
PuTTY Downlo...
Pushpalanka's Blog Hadoop ...
Pushpalanka's Blog Hadoop ...
AWS Marketplace CentOS 6...
Install Hadoop on a Single N...
Technology Inspiration Singl...
Job Vacancies Packages - N...
HadoopExam.com - Adminis...
Find: format
Previous
Next
Options
100%
Start
00:18
13-08-20

There are cryptographic signatures available for all the files we offer below. We also supply cryptographically signed lists of checksums. To download our public keys and find out more about our signature policy, visit the Keys page. If you need a Windows program to compute MD5 checksums, you could try the one at this site. (This MD5 program is also cryptographically signed b its author.)

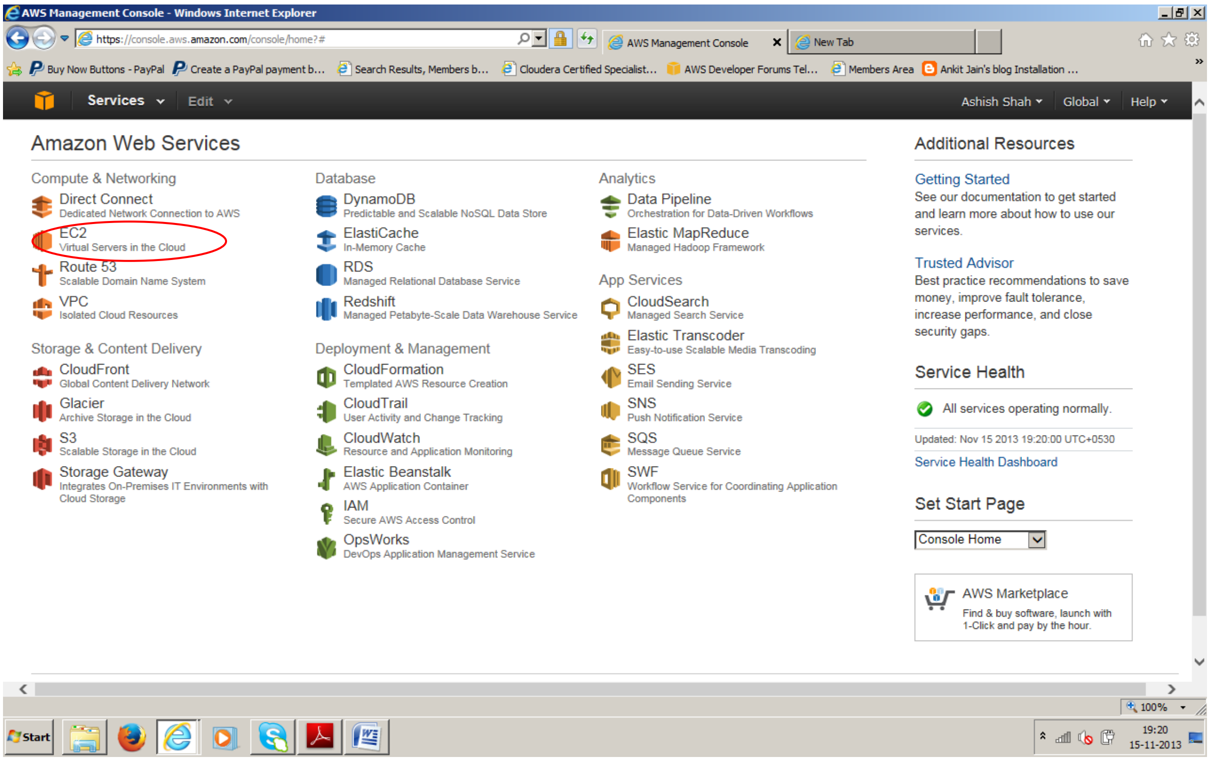
Binaries

The latest release version (beta 0.63). This will generally be a version I think is reasonably likely to work well. If you have a problem with the release version, it might be worth trying out the latest development snapshot (below) to see if I've already fixed the bug, before reporting it to me.

For Windows on Intel x86				
PuTTY:	putty.exe	(or by FTP)	(RSA sig)	(DSA sig)
PuTTYtel:	puttytel.exe	(or by FTP)	(RSA sig)	(DSA sig)
PSCP:	pscp.exe	(or by FTP)	(RSA sig)	(DSA sig)
PSFTP:	psftp.exe	(or by FTP)	(RSA sig)	(DSA sig)
Plink:	plink.exe	(or by FTP)	(RSA sig)	(DSA sig)
Pageant:	pageant.exe	(or by FTP)	(RSA sig)	(DSA sig)
PuTTYgen:	puttygen.exe	(or by FTP)	(RSA sig)	(DSA sig)
A .ZIP file containing all the binaries (except PuTTYtel), and also the help files				
Zip file:	putty.zip	(or by FTP)	(RSA sig)	(DSA sig)
A Windows installer for everything except PuTTYtel				
Installer:	putty-0.63-installer.exe	(or by FTP)	(RSA sig)	(DSA sig)
Checksums for all the above files				
MD5:	md5sums	(or by FTP)	(RSA sig)	(DSA sig)
SHA-1:	sha1sums	(or by FTP)	(RSA sig)	(DSA sig)
SHA-256:	sha256sums	(or by FTP)	(RSA sig)	(DSA sig)
SHA-512:	sha512sums	(or by FTP)	(RSA sig)	(DSA sig)

The latest development snapshot. This will be built every day, automatically, from the current development code - in *whatever* state it's currently in. If you need a fix for a particularly crippling bug, you may well be able to find a fixed PuTTY here well before the fix makes it into the release version above. On the other hand, these snapshots might sometimes be unstable.

1. It requires you have Amazon AWS account. So create/signup Amazon EC2 account by going to http://aws.amazon.com. It also requires you to enter your credit card details. However, it would not charge until you use paid resources. Initially use only micro instances which are free. Click to Amzon EC2 Console under Compute & Networking


AWS Management Console - Windows Internet Explorer
https://console.aws.amazon.com/console/home?#
AWS Management Console
New Tab
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Ankit Jain's blog Installation ...
Services
Edit
Ashish Shah
Global
Help
Amazon Web Services
Compute & Networking
Direct Connect
Dedicated Network Connection to AWS
EC2
Virtual Servers in the Cloud
Route 53
Scalable Domain Name System
VPC
Isolated Cloud Resources
Storage & Content Delivery
CloudFront
Global Content Delivery Network
Glacier
Archive Storage in the Cloud
S3
Scalable Storage in the Cloud
Storage Gateway
Integrates On-Premises IT Environments with Cloud Storage
Database
DynamoDB
Predictable and Scalable NoSQL Data Store
ElastiCache
In-Memory Cache
RDS
Managed Relational Database Service
Redshift
Managed Petabyte-Scale Data Warehouse Service
Deployment & Management
CloudFormation
Templated AWS Resource Creation
CloudTrail
User Activity and Change Tracking
CloudWatch
Resource and Application Monitoring
Elastic Beanstalk
AWS Application Container
IAM
Secure AWS Access Control
OpsWorks
DevOps Application Management Service
Analytics
Data Pipeline
Orchestration for Data-Driven Workflows
Elastic MapReduce
Managed Hadoop Framework
App Services
CloudSearch
Managed Search Service
Elastic Transcoder
Easy-to-use Scalable Media Transcoding
SES
Email Sending Service
SNS
Push Notification Service
SQS
Message Queue Service
SWF
Workflow Service for Coordinating Application Components
Additional Resources
Getting Started
Trusted Advisor
Service Health
Service Health Dashboard
Set Start Page
Console Home
AWS Marketplace
100%
Start
19:20
15-11-2013

See our documentation to get started and learn more about how to use our services.

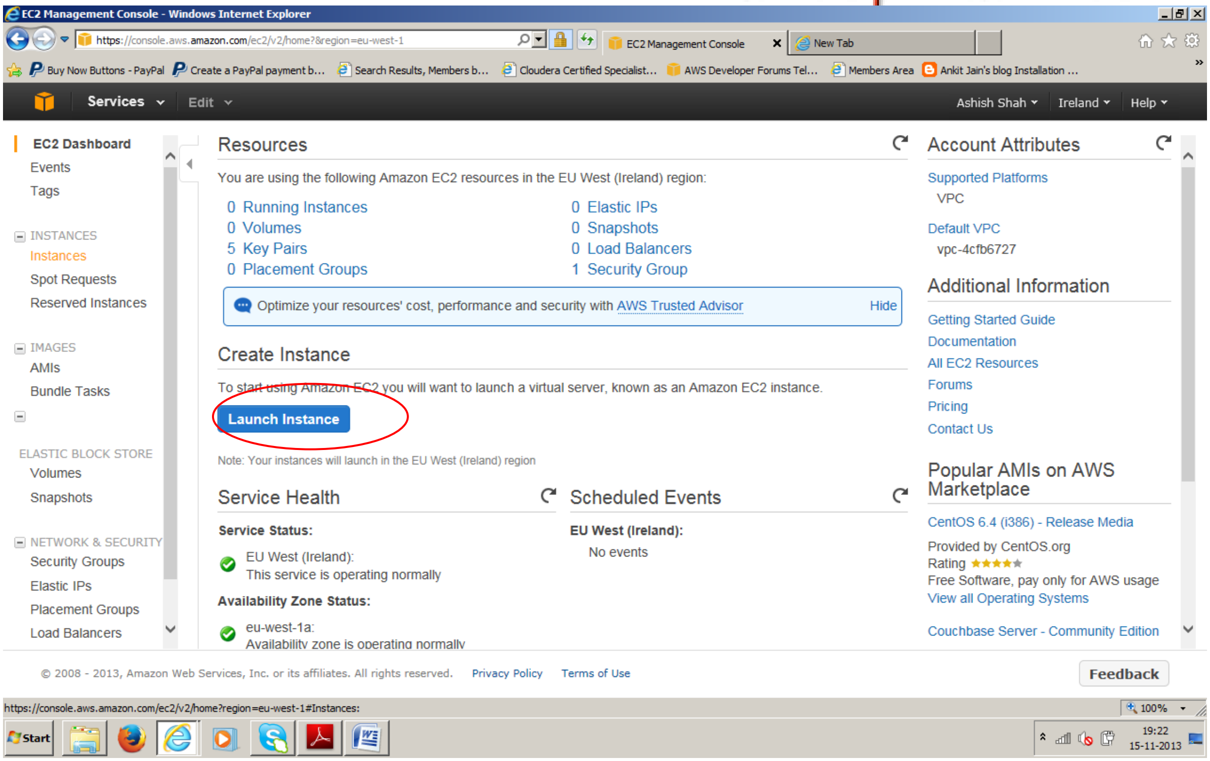
Best practice recommendations to save money, improve fault tolerance, increase performance, and close security gaps.

All services operating normally.

Updated: Nov 15 2013 19:20:00 UTC+0530

Find & buy software, launch with 1-Click and pay by the hour.

2. Once you are in then Click Launch Instance (chose EU West Ireland region). This will create a Virtual Machine Instance in the cloud. And you have to provide the configuration which you can see in next steps



Resources

You are using the following Amazon EC2 resources in the EU West (Ireland) region:

0 Running Instances
0 Volumes
5 Key Pairs
0 Placement Groups
0 Elastic IPs
0 Snapshots
0 Load Balancers
1 Security Group

Optimize your resources' cost, performance and security with AWS Trusted Advisor

Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

Note: Your instances will launch in the EU West (Ireland) region

Service Health

Service Status:

EU West (Ireland):
This service is operating normally

Availability Zone Status:

eu-west-1a:
Availability zone is operating normally

Scheduled Events

EU West (Ireland):
No events

Account Attributes

Supported Platforms
VPC

Default VPC
vpc-4cfb6727

Additional Information

Getting Started Guide
Documentation
All EC2 Resources
Forums
Pricing
Contact Us

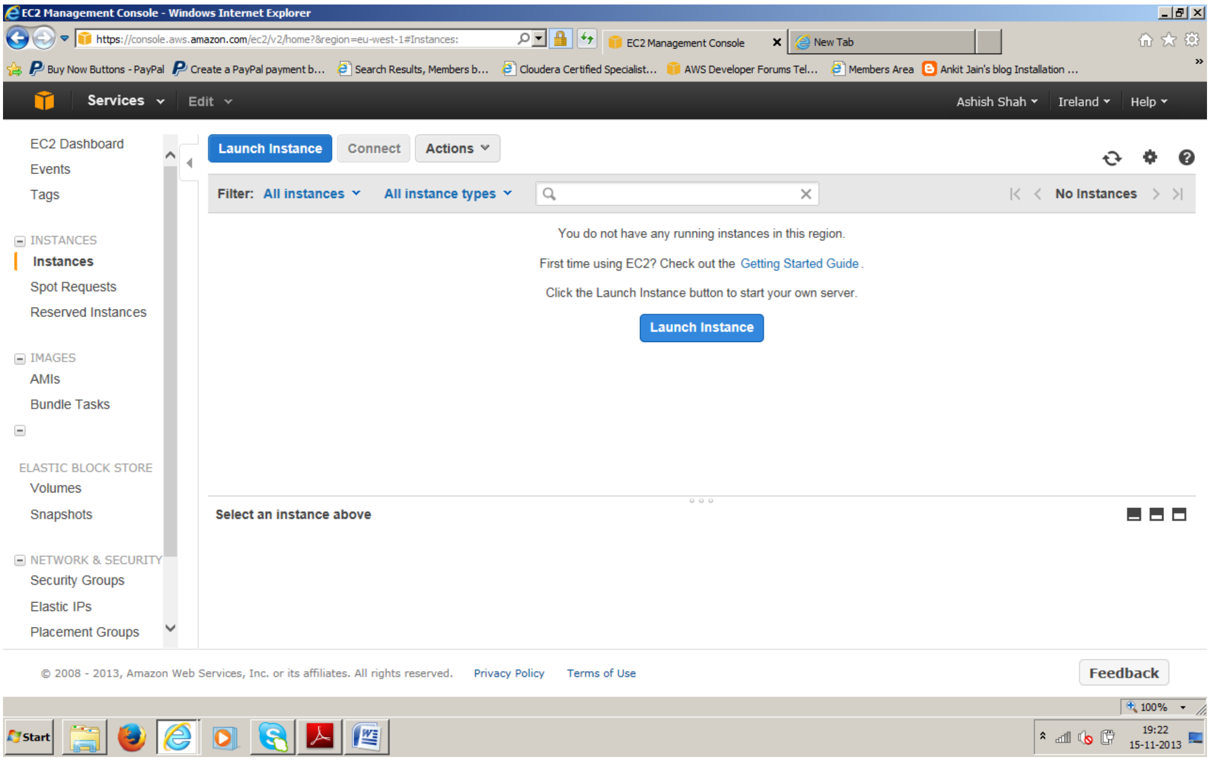
Popular AMIs on AWS Marketplace

CentOS 6.4 (i386) - Release Media
Provided by CentOS.org
Rating ★★★★★
Free Software, pay only for AWS usage
View all Operating Systems

Couchbase Server - Community Edition

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved.

https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#Instances:



You do not have any running instances in this region.

First time using EC2? Check out the Getting Started Guide.

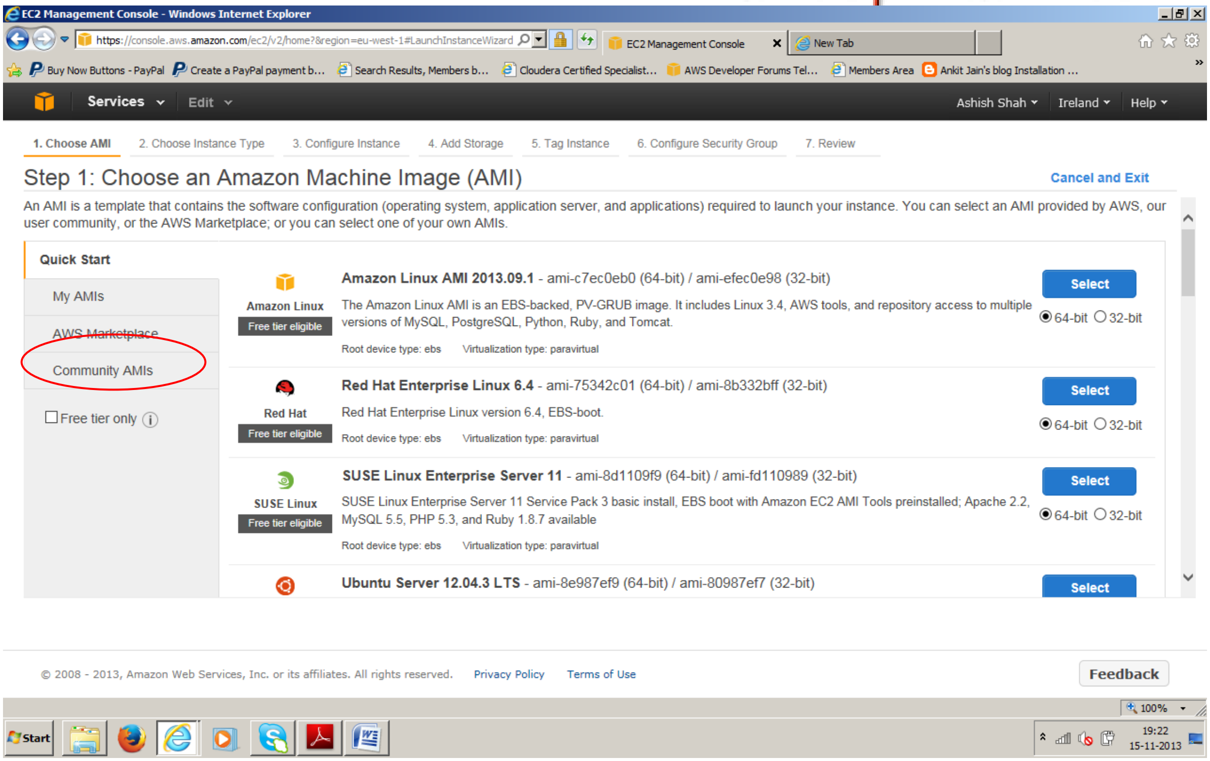
Click the Launch Instance button to start your own server.

Select an instance above

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved.

3. Select Community AMI's

An Amazon Machine Image (AMI) is a special type of pre-configured operating system and virtual application software which is used to create a virtual machine within the Amazon Elastic Compute Cloud (EC2). It serves as the basic unit of deployment for services delivered using EC2.


EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?&region=eu-west-1#LaunchInstanceWizard
EC2 Management Console
New Tab
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Ankit Jain's blog Installation ...
Services
Edit
Ashish Shah
Ireland
Help
1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Tag Instance
6. Configure Security Group
7. Review
Cancel and Exit
Quick Start
My AMIs
AWS Marketplace
Community AMIs
Select
Feedback
Start

Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

□ Free tier only

Amazon Linux — Free tier eligible
Amazon Linux AMI 2013.09.1 - ami-c7ec0eb0 (64-bit) / ami-efec0e98 (32-bit)
The Amazon Linux AMI is an EBS-backed, PV-GRUB image. It includes Linux 3.4, AWS tools, and repository access to multiple versions of MySQL, PostgreSQL, Python, Ruby, and Tomcat.
Root device type: ebs Virtualization type: paravirtual
● 64-bit ○ 32-bit

Red Hat — Free tier eligible
Red Hat Enterprise Linux 6.4 - ami-75342c01 (64-bit) / ami-8b332bff (32-bit)
Red Hat Enterprise Linux version 6.4, EBS-boot.
Root device type: ebs Virtualization type: paravirtual
● 64-bit ○ 32-bit

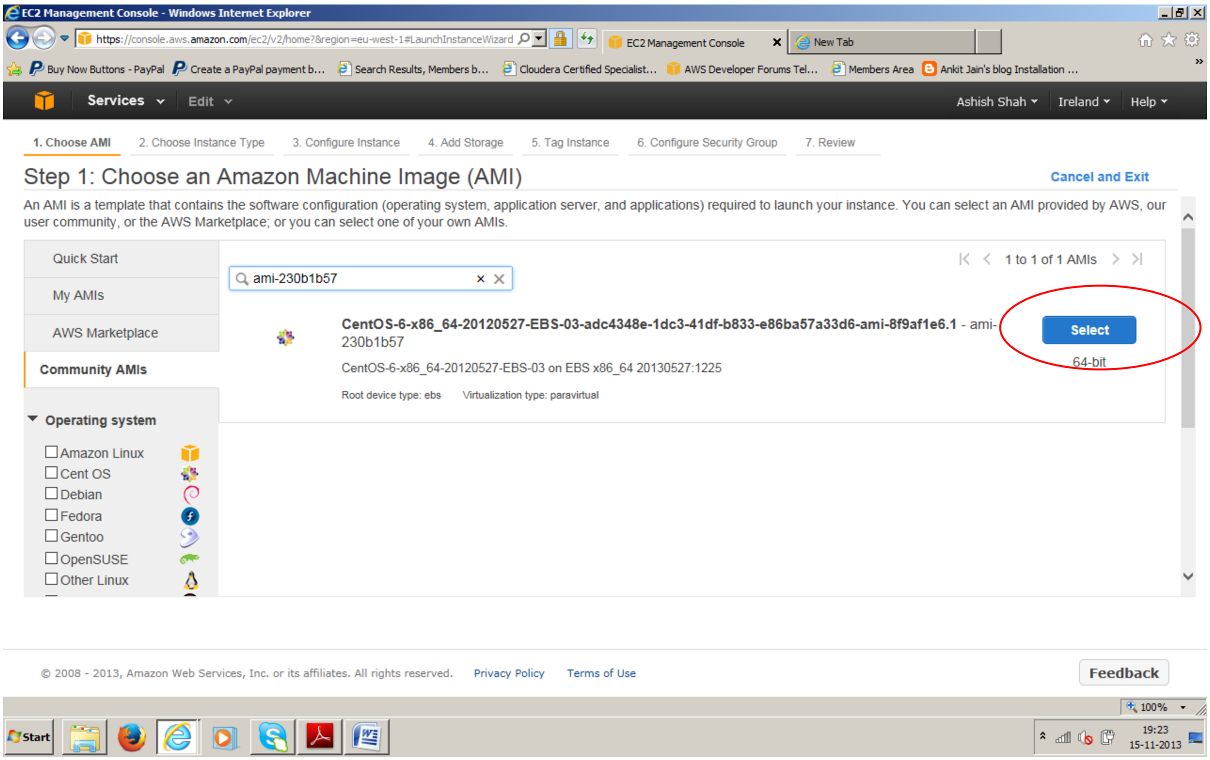
SUSE Linux — Free tier eligible
SUSE Linux Enterprise Server 11 - ami-8d1109f9 (64-bit) / ami-fd110989 (32-bit)
SUSE Linux Enterprise Server 11 Service Pack 3 basic install, EBS boot with Amazon EC2 AMI Tools preinstalled; Apache 2.2, MySQL 5.5, PHP 5.3, and Ruby 1.8.7 available
Root device type: ebs Virtualization type: paravirtual
● 64-bit ○ 32-bit

Ubuntu Server 12.04.3 LTS - ami-8e987ef9 (64-bit) / ami-80987ef7 (32-bit)

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

100%
19:22
15-11-2013

4. Search for Community AMI: Now there are lot of pre-configured AMIs available in Amazon EC2 cloud. You can search in AWS Marketplace as well. We are choosing AMI's for CentOS linux 6.0 version and id for this is (**ami-230b1b57**)


EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?&region=eu-west-1#LaunchInstanceWizard
EC2 Management Console
New Tab
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Ankit Jain's blog Installation ...
Services
Edit
Ashish Shah
Ireland
Help
1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Tag Instance
6. Configure Security Group
7. Review
Cancel and Exit
Quick Start
My AMIs
AWS Marketplace
Community AMIs
Select
Feedback
Start

Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

ami-230b1b57

1 to 1 of 1 AMIs

CentOS-6-x86_64-20120527-EBS-03-adc4348e-1dc3-41df-b833-e86ba57a33d6-ami-8f9af1e6.1 - ami-230b1b57
CentOS-6-x86_64-20120527-EBS-03 on EBS x86_64 20130527:1225
Root device type: ebs Virtualization type: paravirtual
64-bit

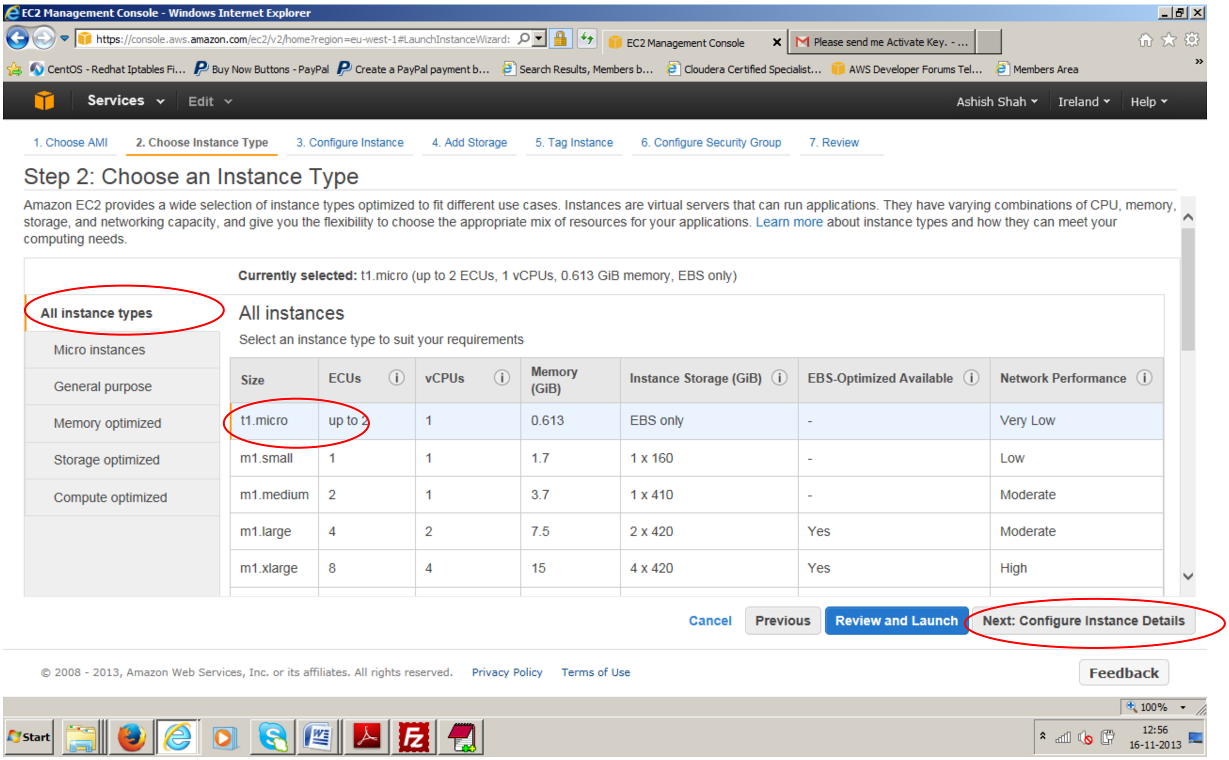
Operating system
□ Amazon Linux
□ Cent OS
□ Debian
□ Fedora
□ Gentoo
□ OpenSUSE
□ Other Linux

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

100%
19:23
15-11-2013

5. Now in this step we will decide how many instance of this virtual machine and type of the instance. We are going to create multi node cluster hence select 3 instance and choose **Small** Instance type which at least required for running Hadoop mapreduce example. You

can choose micro instance which is completely free for 750 Hrs in a month, but that is not enough to run mapred example. However, if you are new to EC2 we suggest you try with micro instance first, so you would not incur any cost while configuring Hadoop cluster. And once you become confident with the configuration then you can start using the Small Instance for real practice. However, cost is very small approx.06$(Check Amazon for price) per Hour per Instance. And now select on m1.small and then click Next: Configure Instance Details.


EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard:
EC2 Management Console
Please send me Activate Key. - ...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Services
Edit
Ashish Shah
Ireland
Help
1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Tag Instance
6. Configure Security Group
7. Review

Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn more about instance types and how they can meet your computing needs.

Currently selected: t1.micro (up to 2 ECUs, 1 vCPUs, 0.613 GiB memory, EBS only)

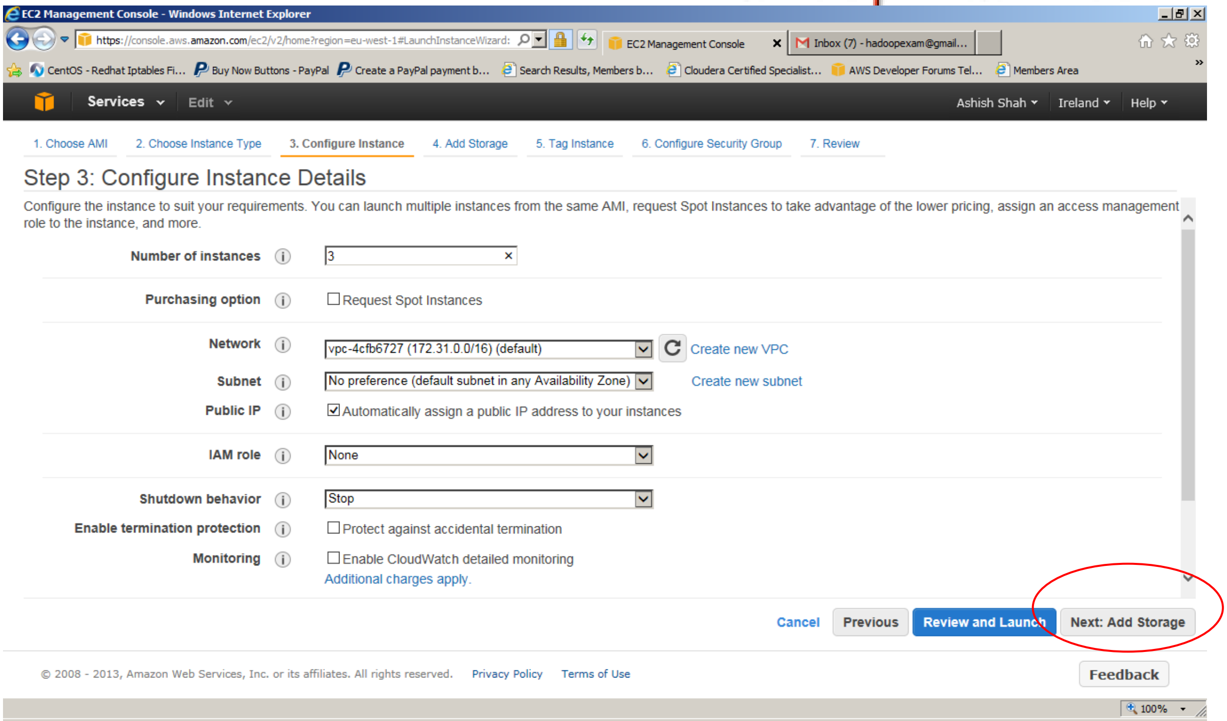
All instances

Select an instance type to suit your requirements

Size	ECUs	vCPUs	Memory (GiB)	Instance Storage (GiB)	EBS-Optimized Available	Network Performance
t1.micro	up to 2	1	0.613	EBS only	-	Very Low
m1.small	1	1	1.7	1 x 160	-	Low
m1.medium	2	1	3.7	1 x 410	-	Moderate
m1.large	4	2	7.5	2 x 420	Yes	Moderate
m1.xlarge	8	4	15	4 x 420	Yes	High

Cancel
Previous
Review and Launch
Next: Configure Instance Details
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Feedback
100%
Start
12:56
16-11-2013

6. Type Number of Instances to 3 (We will create 3 node cluster 1 master and 2 slave nodes)



EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard:
EC2 Management Console
Inbox (7) - hadoopexam@gmail...
CentOS - Redhat Iptables Fi... | Buy Now Buttons - PayPal | Create a PayPal payment b... | Search Results, Members b... | Cloudera Certified Specialist... | AWS Developer Forums Tel... | Members Area
Services | Edit | Ashish Shah | Ireland | Help
1. Choose AMI | 2. Choose Instance Type | 3. Configure Instance | 4. Add Storage | 5. Tag Instance | 6. Configure Security Group | 7. Review

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot Instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of instances: 3

Purchasing option: □ Request Spot Instances

Network: vpc-4cfb6727 (172.31.0.0/16) (default) — Create new VPC

Subnet: No preference (default subnet in any Availability Zone) — Create new subnet

Public IP: ☑ Automatically assign a public IP address to your instances

IAM role: None

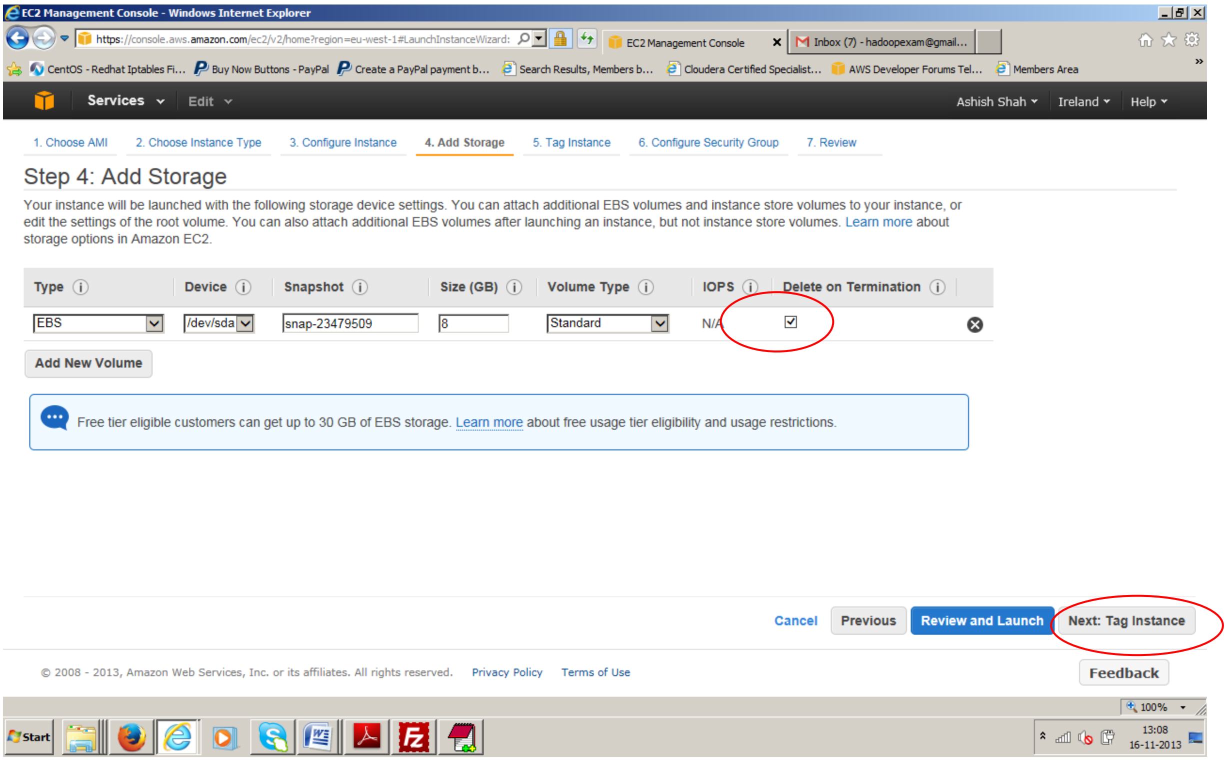
Shutdown behavior: Stop

Enable termination protection: □ Protect against accidental termination

Monitoring: □ Enable CloudWatch detailed monitoring
Additional charges apply.

Cancel | Previous | Review and Launch | Next: Add Storage
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use
Feedback
100%

7. We also need to attach storage to all instances as you know we need storage space as well. Keep in mind you check the “Delete On Termination” as below. So as soon as you terminate your instances all attached storage will be deleted and you avoid storage charge.



EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard:
EC2 Management Console
Inbox (7) - hadoopexam@gmail...
CentOS - Redhat Iptables Fi... | Buy Now Buttons - PayPal | Create a PayPal payment b... | Search Results, Members b... | Cloudera Certified Specialist... | AWS Developer Forums Tel... | Members Area
Services | Edit | Ashish Shah | Ireland | Help
1. Choose AMI | 2. Choose Instance Type | 3. Configure Instance | 4. Add Storage | 5. Tag Instance | 6. Configure Security Group | 7. Review

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. Learn more about storage options in Amazon EC2.

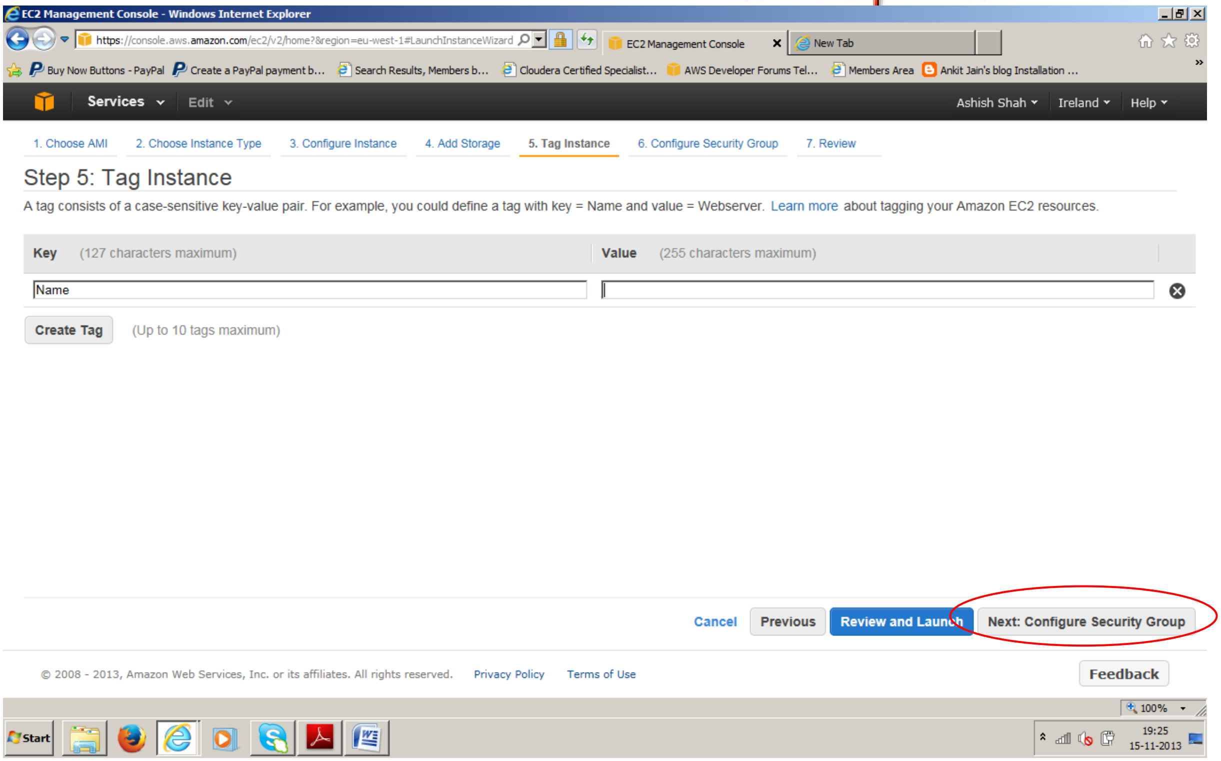
Type	Device	Snapshot	Size (GB)	Volume Type	IOPS	Delete on Termination
EBS	/dev/sda	snap-23479509	8	Standard	N/A	☑

Add New Volume

Free tier eligible customers can get up to 30 GB of EBS storage. Learn more about free usage tier eligibility and usage restrictions.

Cancel | Previous | Review and Launch | Next: Tag Instance
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use
Feedback
100%
Start
13:08
16-11-2013

8. Tag Instances : We will tag to instances later on, just click on “Configure Security Group” as below.


EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?&region=eu-west-1#LaunchInstanceWizard
Services
Edit
Ashish Shah
Ireland
Help
1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Tag Instance
6. Configure Security Group
7. Review
Create Tag
Cancel
Previous
Review and Launch
Next: Configure Security Group
Feedback

Step 5: Tag Instance

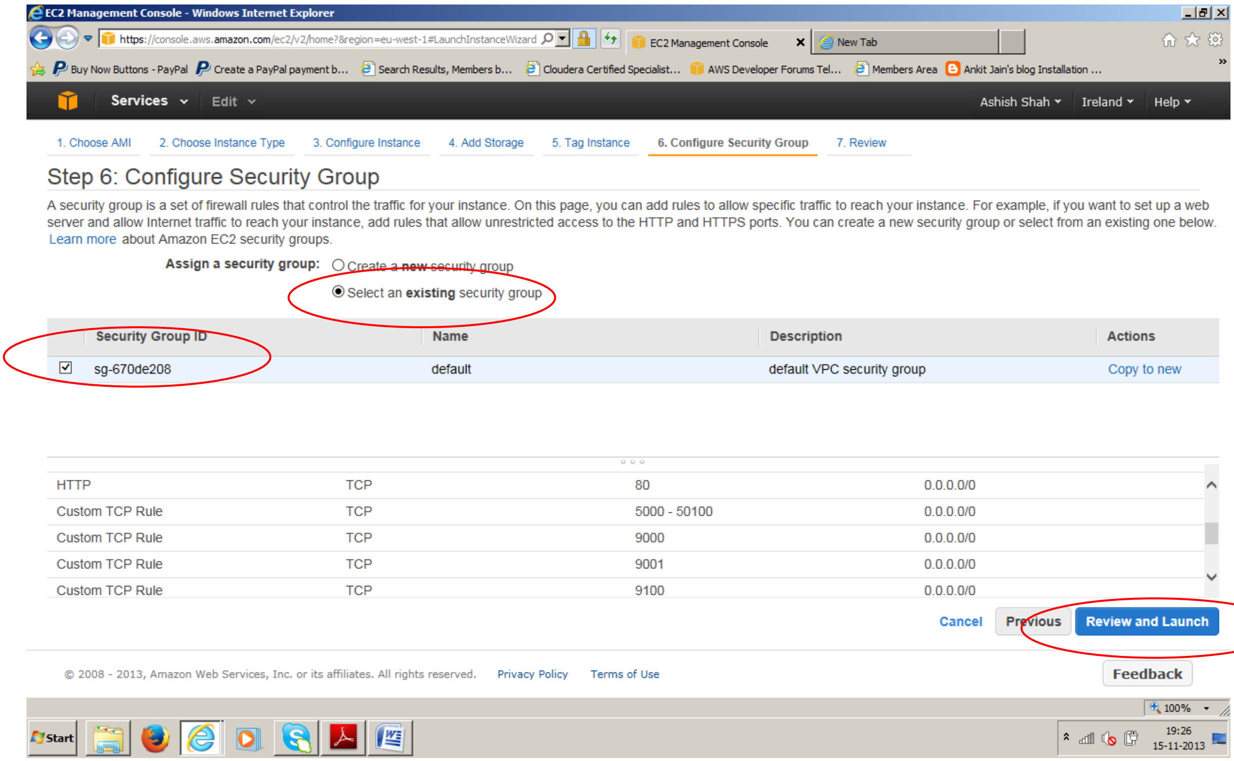
A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. Learn more about tagging your Amazon EC2 resources.

Key (127 characters maximum)	Value (255 characters maximum)
Name	

(Up to 10 tags maximum)

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

9. Select the existing Security group, as we already created in the Module 14 training. If you have not done yet than watch Module 14 first. And then click Review and Launch.



Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. Learn more about Amazon EC2 security groups.

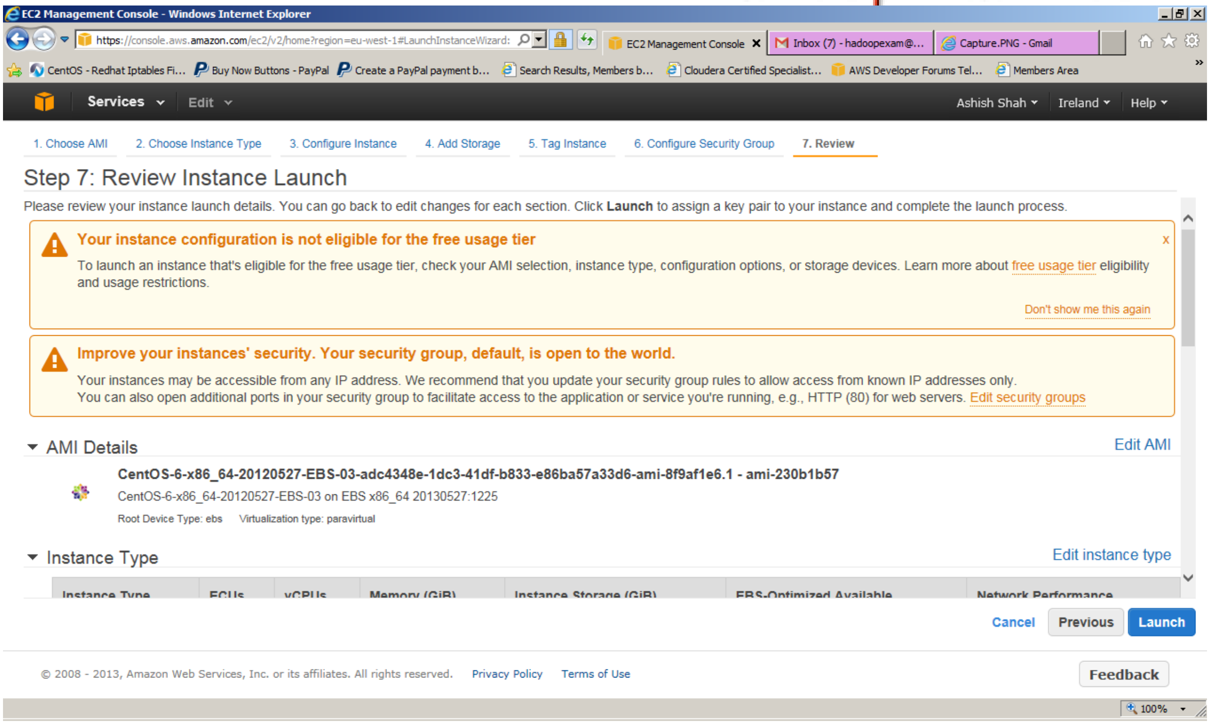
Assign a security group: ○ Create a new security group
● Select an existing security group

	Security Group ID	Name	Description	Actions
☑	sg-670de208	default	default VPC security group	Copy to new

HTTP	TCP	80	0.0.0.0/0
Custom TCP Rule	TCP	5000 - 50100	0.0.0.0/0
Custom TCP Rule	TCP	9000	0.0.0.0/0
Custom TCP Rule	TCP	9001	0.0.0.0/0
Custom TCP Rule	TCP	9100	0.0.0.0/0

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

10. Review your configuration in this screen, ignore these warnings and click launch after that



EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard:
Services
Edit
Ashish Shah
Ireland
Help

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

Your instance configuration is not eligible for the free usage tier

To launch an instance that's eligible for the free usage tier, check your AMI selection, instance type, configuration options, or storage devices. Learn more about free usage tier eligibility and usage restrictions.

Don't show me this again

Improve your instances' security. Your security group, default, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.
You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. Edit security groups

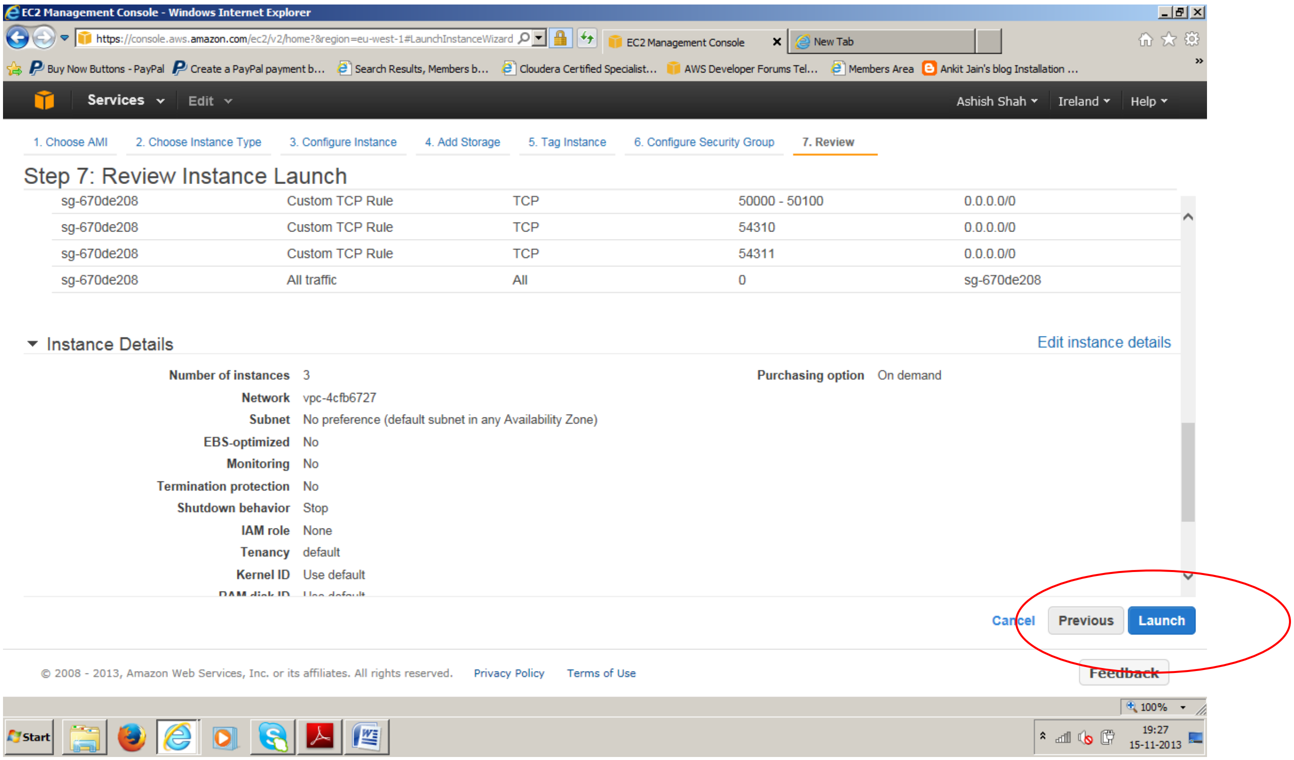
AMI Details — Edit AMI

CentOS-6-x86_64-20120527-EBS-03-adc4348e-1dc3-41df-b833-e86ba57a33d6-ami-8f9af1e6.1 - ami-230b1b57
CentOS-6-x86_64-20120527-EBS-03 on EBS x86_64 20130527:1225
Root Device Type: ebs Virtualization type: paravirtual

Instance Type — Edit instance type

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GiB)	EBS-Optimized Available	Network Performance

Cancel
Previous
Launch
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
Feedback



EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard
Services
Edit
Ashish Shah
Ireland
Help

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

Step 7: Review Instance Launch

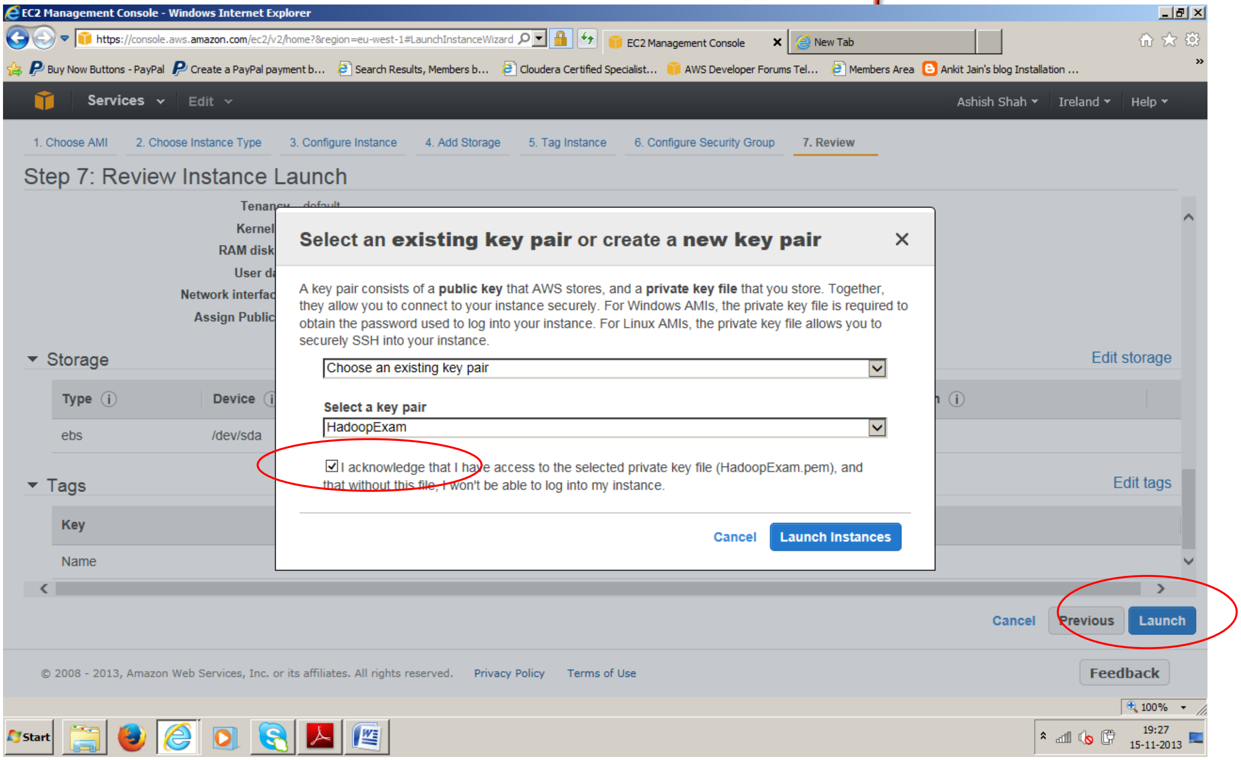
sg-670de208	Custom TCP Rule	TCP	50000 - 50100	0.0.0.0/0
sg-670de208	Custom TCP Rule	TCP	54310	0.0.0.0/0
sg-670de208	Custom TCP Rule	TCP	54311	0.0.0.0/0
sg-670de208	All traffic	All	0	sg-670de208

Instance Details — Edit instance details

Number of instances	3	Purchasing option	On demand
Network	vpc-4cfb6727		
Subnet	No preference (default subnet in any Availability Zone)		
EBS-optimized	No		
Monitoring	No		
Termination protection	No		
Shutdown behavior	Stop		
IAM role	None		
Tenancy	default		
Kernel ID	Use default		
RAM disk ID	Use default		

Cancel
Previous
Launch
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
Feedback
Start
19:27
15-11-2013

11. To connect to cluster we need private key. So please select the existing key which we have created in module 14. And check the acknowledge box.



Step 7: Review Instance Launch

Storage

Type	Device
ebs	/dev/sda

Tags

Key
Name

Select an existing key pair or create a new key pair

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Choose an existing key pair

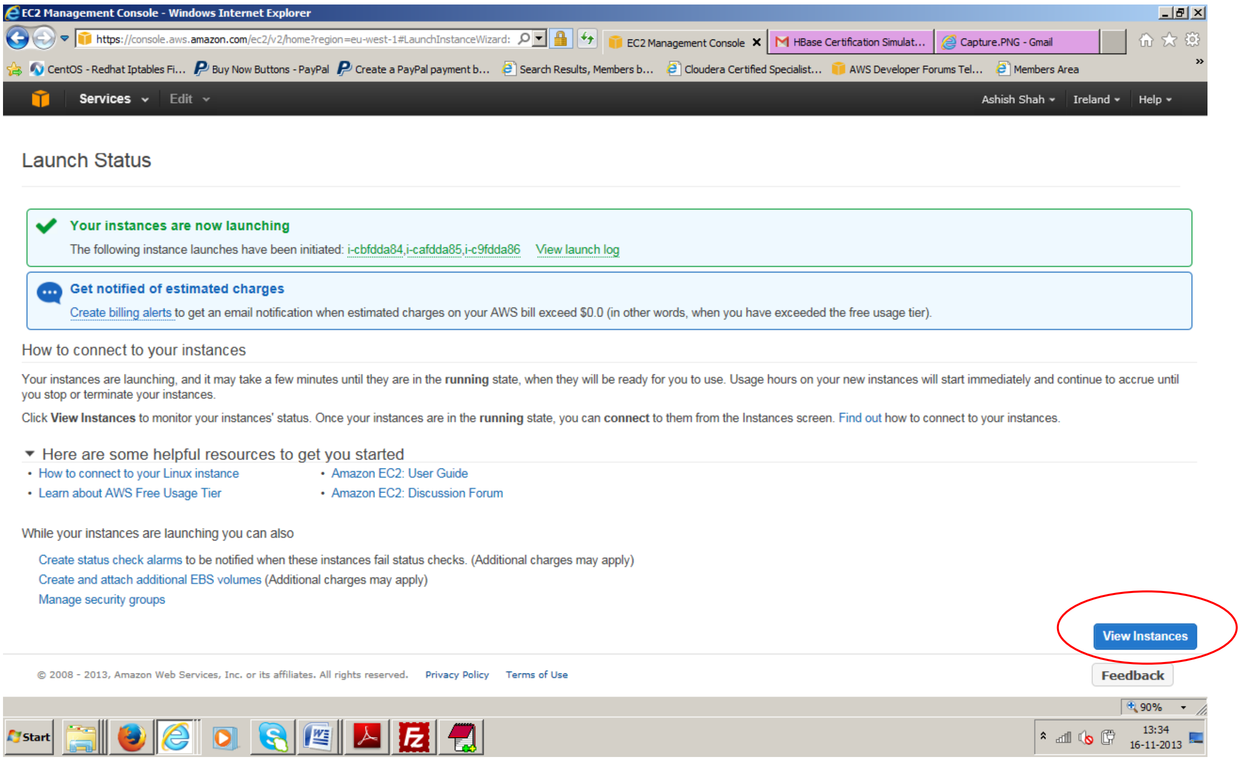
Select a key pair
HadoopExam

☑ I acknowledge that I have access to the selected private key file (HadoopExam.pem), and that without this file, I won't be able to log into my instance.

Cancel | Launch Instances

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

12. Launch Instances status as below, now click View Instances.


EC2 Management Console - Windows Internet Explorer
https://console.aws.amazon.com/ec2/v2/home?region=eu-west-1#LaunchInstanceWizard:
Services
Edit
Ashish Shah
Ireland
Help
View Instances
Feedback

Launch Status

Your instances are now launching
The following instance launches have been initiated: i-cbfdda84,i-cafdda85,i-c9fdda86 View launch log

Get notified of estimated charges
Create billing alerts to get an email notification when estimated charges on your AWS bill exceed $0.0 (in other words, when you have exceeded the free usage tier).

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances.

Click **View Instances** to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the Instances screen. Find out how to connect to your instances.

Here are some helpful resources to get you started

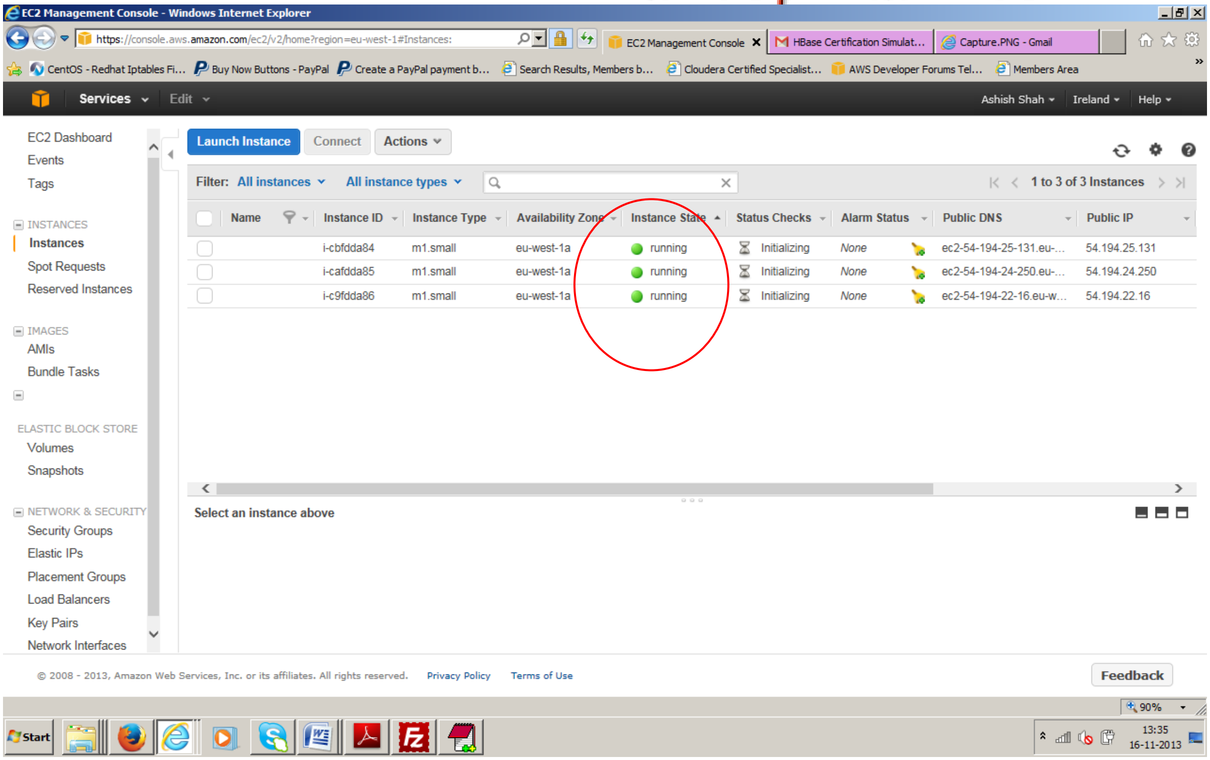
- How to connect to your Linux instance
- Learn about AWS Free Usage Tier
- Amazon EC2: User Guide
- Amazon EC2: Discussion Forum

While your instances are launching you can also

Create status check alarms to be notified when these instances fail status checks. (Additional charges may apply)
Create and attach additional EBS volumes (Additional charges may apply)
Manage security groups

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

13. You should be able to see three running instances as below.

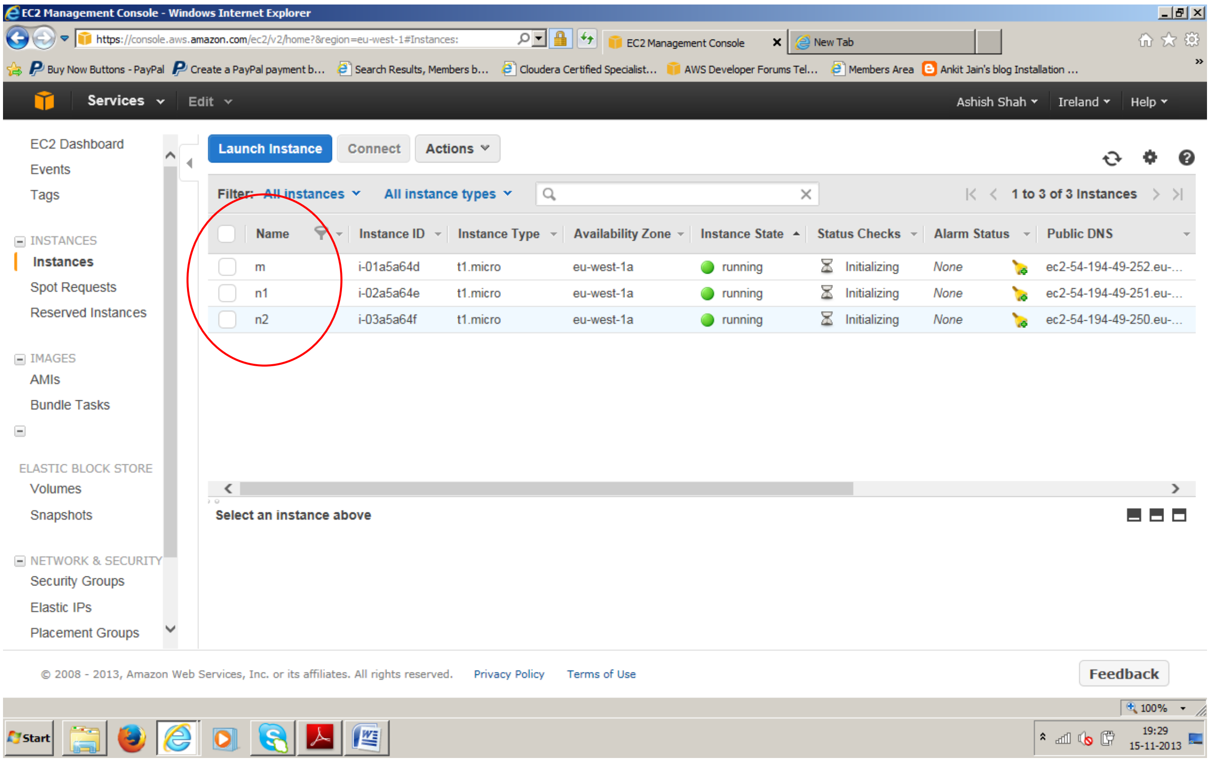


Filter: All instances · All instance types — 1 to 3 of 3 Instances

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
	i-cbfdda84	m1.small	eu-west-1a	running	Initializing	None	ec2-54-194-25-131.eu-...	54.194.25.131
	i-cafdda85	m1.small	eu-west-1a	running	Initializing	None	ec2-54-194-24-250.eu-...	54.194.24.250
	i-c9fdda86	m1.small	eu-west-1a	running	Initializing	None	ec2-54-194-22-16.eu-w...	54.194.22.16

Select an instance above

14. Now rename the Instances as (m:master, n1 and n2 as datanode) as below.

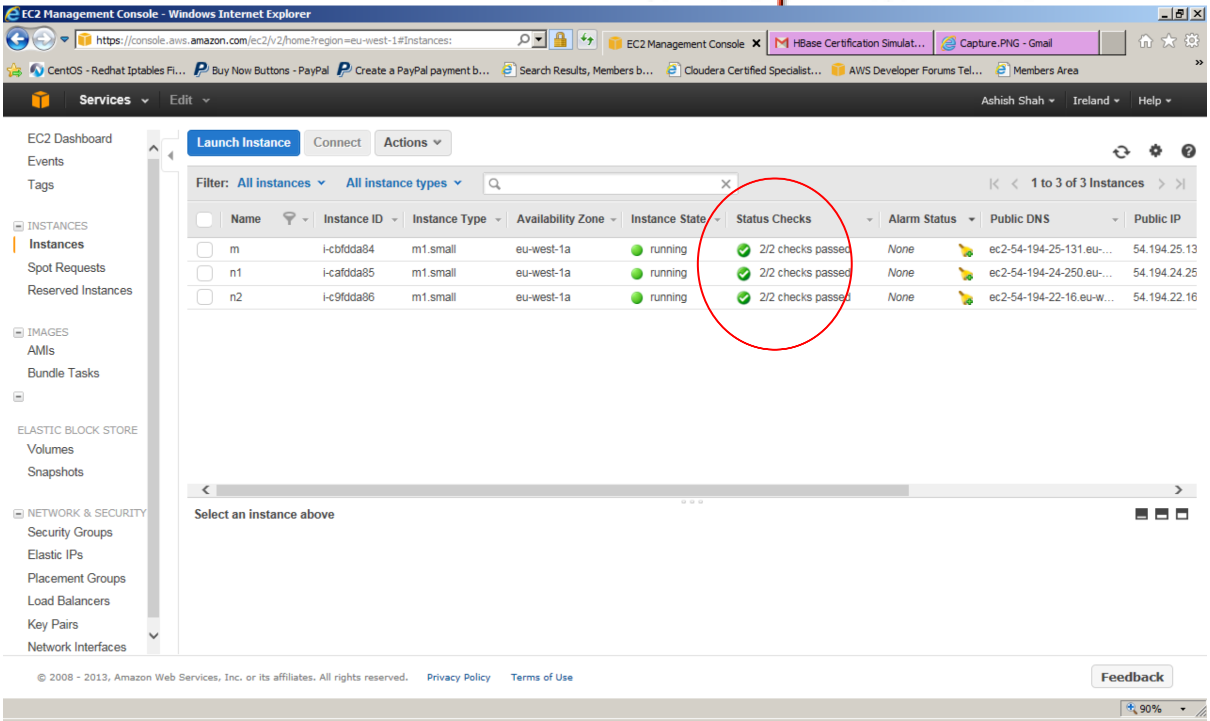


Filter: All instances · All instance types — 1 to 3 of 3 Instances

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS
m	i-01a5a64d	t1.micro	eu-west-1a	running	Initializing	None	ec2-54-194-49-252.eu-...
n1	i-02a5a64e	t1.micro	eu-west-1a	running	Initializing	None	ec2-54-194-49-251.eu-...
n2	i-03a5a64f	t1.micro	eu-west-1a	running	Initializing	None	ec2-54-194-49-250.eu-...

Select an instance above

15. Wait status change from Initializing to “2/2 checks passed” as below



Filter: All instances · All instance types · 1 to 3 of 3 Instances

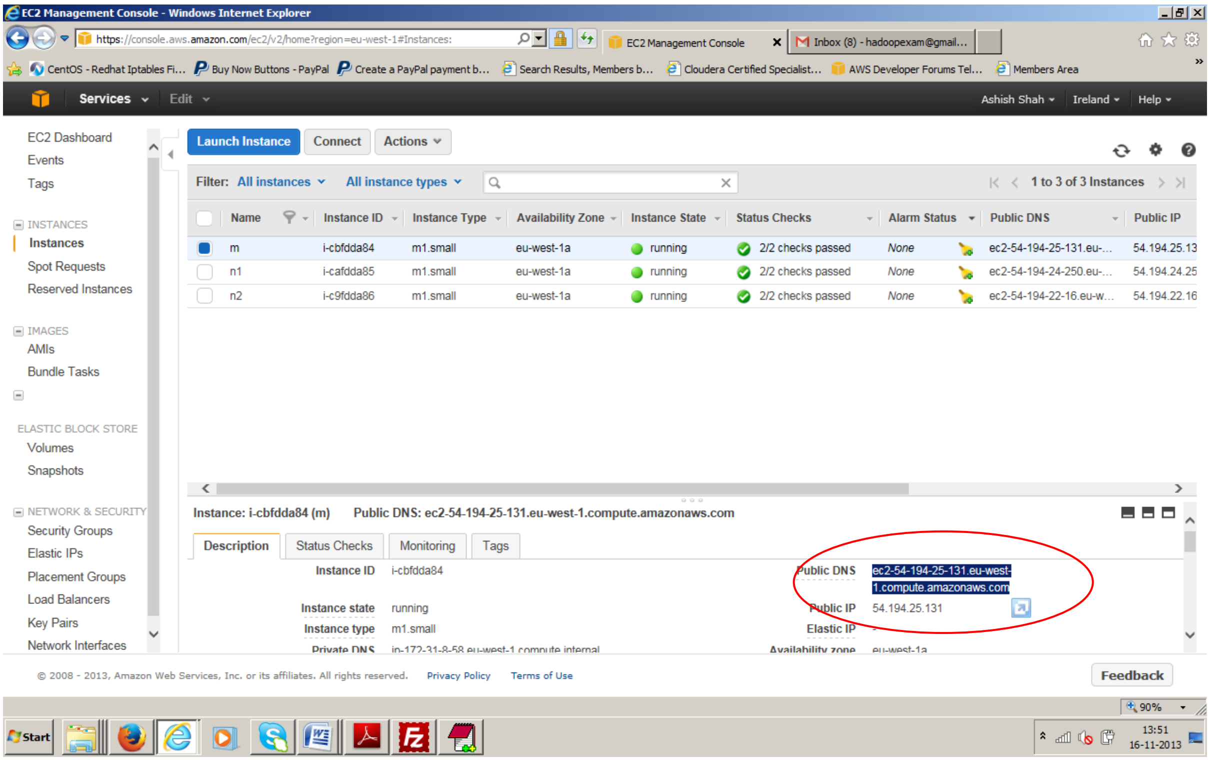
Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25.13
n1	i-cafdda85	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24.25
n2	i-c9fdda86	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-22-16.eu-w...	54.194.22.16

Select an instance above

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

16. Copy the Public DNS of all the nodes somewhere in notepad for future use and this is URL by which you will access your instance which you have created as below, using the putty.

Public DNS of Master as below (It would be different for each new instances)



Filter: All instances · All instance types · 1 to 3 of 3 Instances

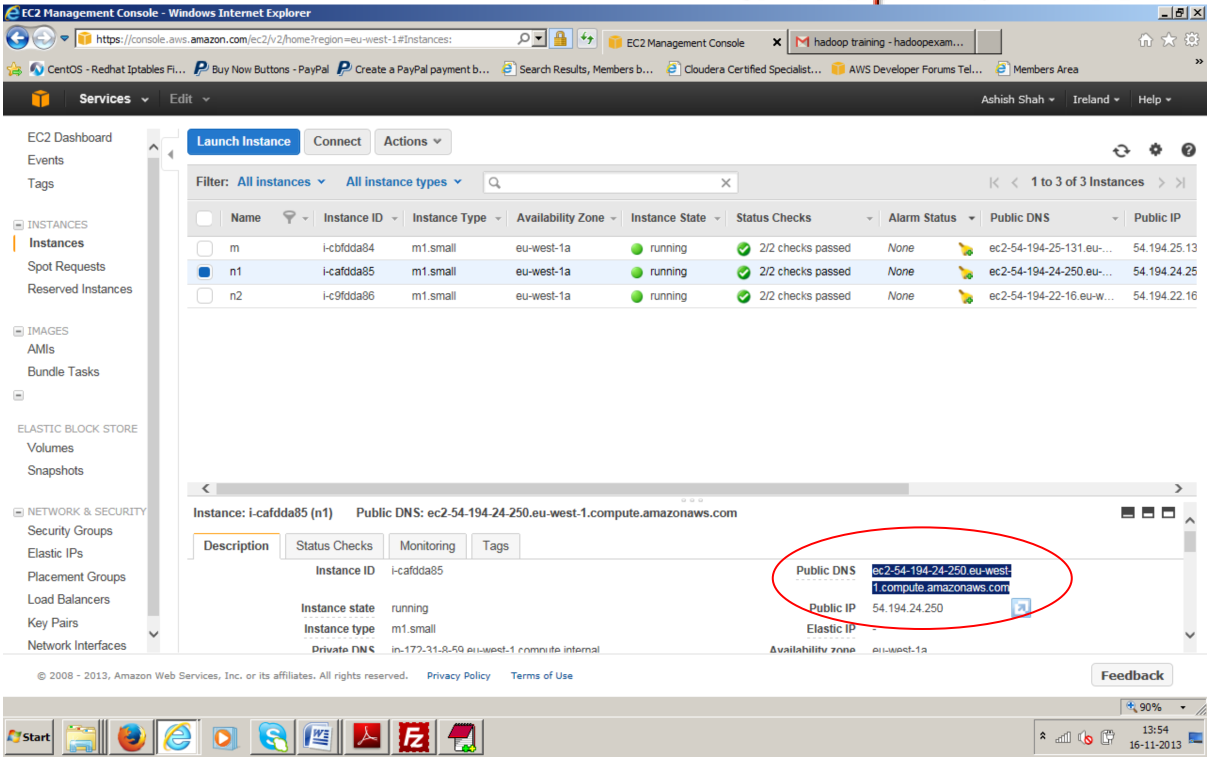
Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25.13
n1	i-cafdda85	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24.25
n2	i-c9fdda86	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-22-16.eu-w...	54.194.22.16

Instance: i-cbfdda84 (m) Public DNS: ec2-54-194-25-131.eu-west-1.compute.amazonaws.com

Instance ID	i-cbfdda84	Public DNS	ec2-54-194-25-131.eu-west-1.compute.amazonaws.com
Instance state	running	Public IP	54.194.25.131
Instance type	m1.small	Elastic IP	-
Private DNS	ip-172-31-8-58.eu-west-1.compute.internal	Availability zone	eu-west-1a

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Public DNS of DataNode n1



Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25.13
n1	i-cafdda85	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24.25
n2	i-c9fdda86	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-22-16.eu-w...	54.194.22.16

Instance: i-cafdda85 (n1) Public DNS: ec2-54-194-24-250.eu-west-1.compute.amazonaws.com

Instance ID i-cafdda85

Public DNS ec2-54-194-24-250.eu-west-1.compute.amazonaws.com

Instance state running

Public IP 54.194.24.250

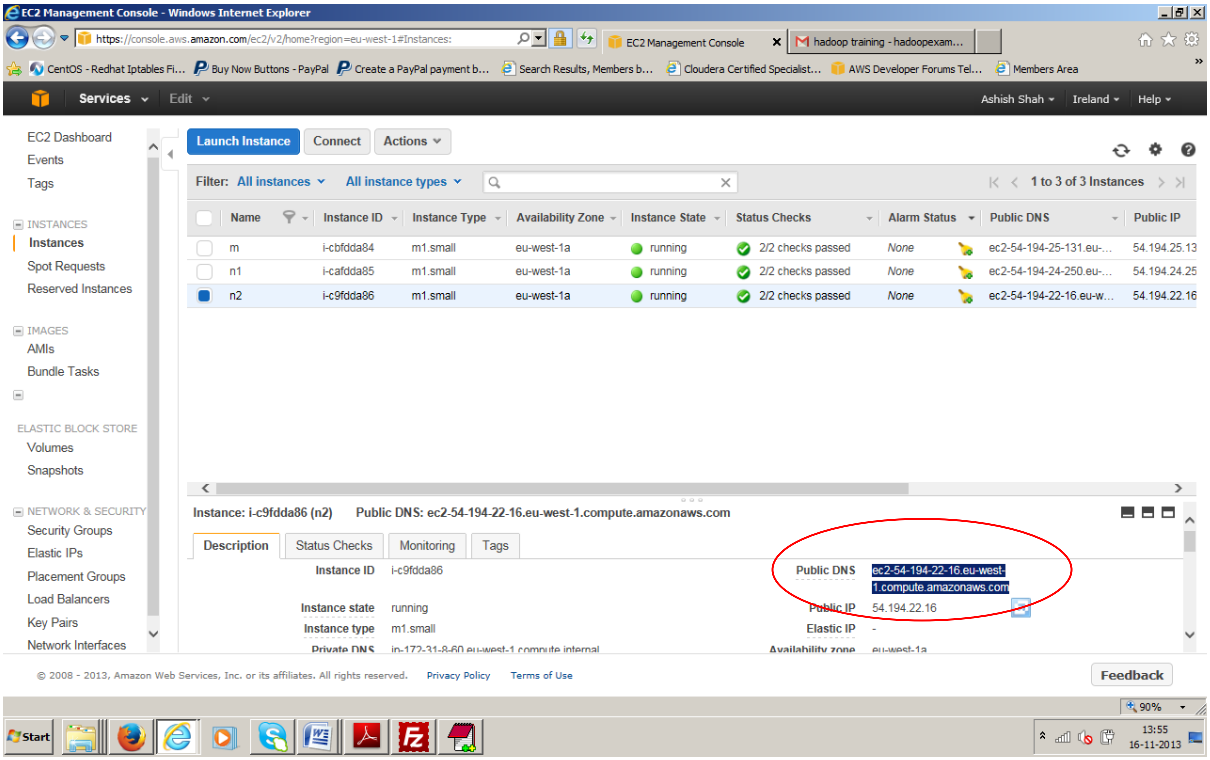
Instance type m1.small

Elastic IP -

Private DNS ip-172-31-8-59.eu-west-1.compute.internal

Availability zone eu-west-1a

Public DNS of DataNode n2



Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25.13
n1	i-cafdda85	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24.25
n2	i-c9fdda86	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-22-16.eu-w...	54.194.22.16

Instance: i-c9fdda86 (n2) Public DNS: ec2-54-194-22-16.eu-west-1.compute.amazonaws.com

Instance ID i-c9fdda86

Public DNS ec2-54-194-22-16.eu-west-1.compute.amazonaws.com

Instance state running

Public IP 54.194.22.16

Instance type m1.small

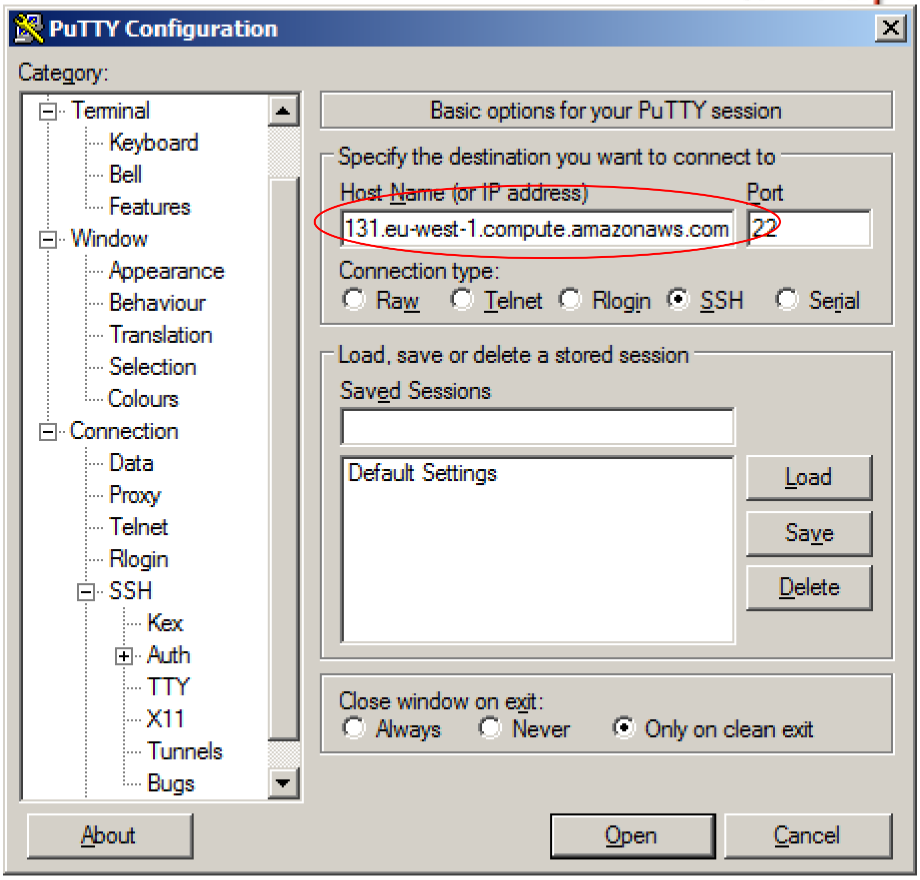
Elastic IP -

Private DNS ip-172-31-8-60.eu-west-1.compute.internal

Availability zone eu-west-1a

17. Now connect to all nodes using putty as below (In module 14 we have shown Putty and Other tool use).

Put the HostName(Public DNS) of Master



PuTTY Configuration

Basic options for your PuTTY session

Specify the destination you want to connect to

Host Name (or IP address): 131.eu-west-1.compute.amazonaws.com

Port: 22

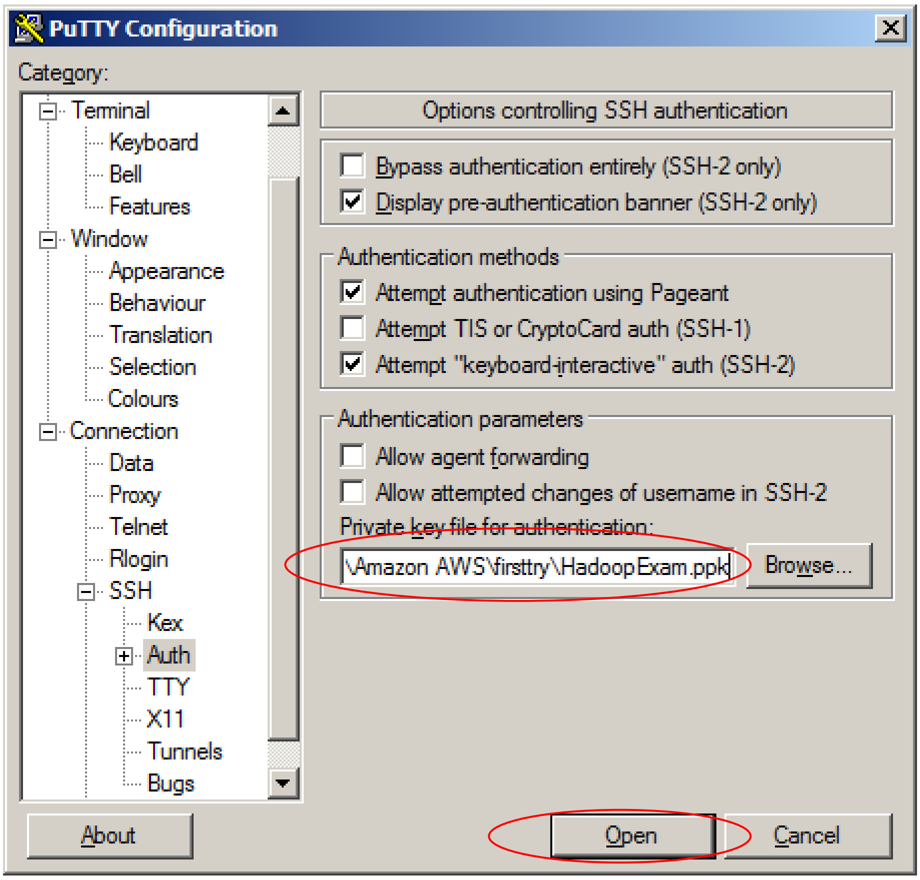
Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Load, save or delete a stored session

Saved Sessions: Default Settings

Close window on exit: ○ Always ○ Never ● Only on clean exit

Select the Private Key to connect this node and click open.



PuTTY Configuration

Options controlling SSH authentication

□ Bypass authentication entirely (SSH-2 only)
☑ Display pre-authentication banner (SSH-2 only)

Authentication methods

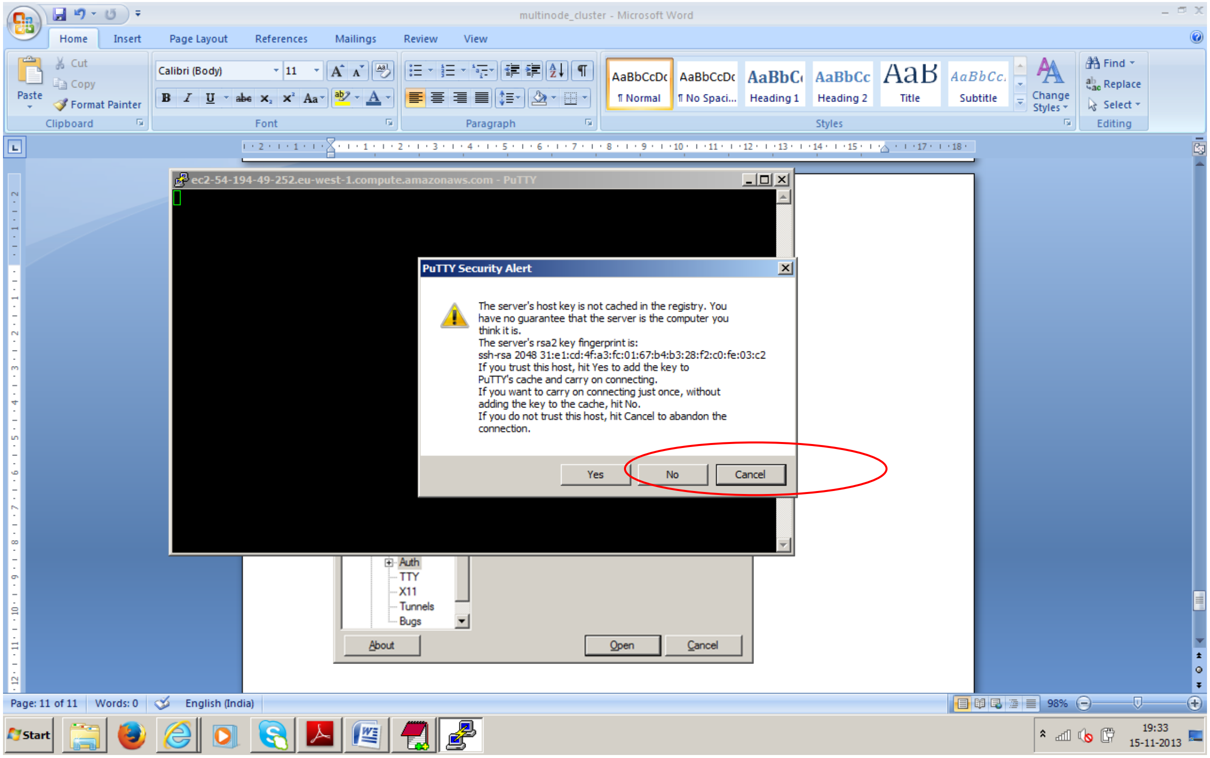
☑ Attempt authentication using Pageant
□ Attempt TIS or CryptoCard auth (SSH-1)
☑ Attempt "keyboard-interactive" auth (SSH-2)

Authentication parameters

□ Allow agent forwarding
□ Allow attempted changes of username in SSH-2

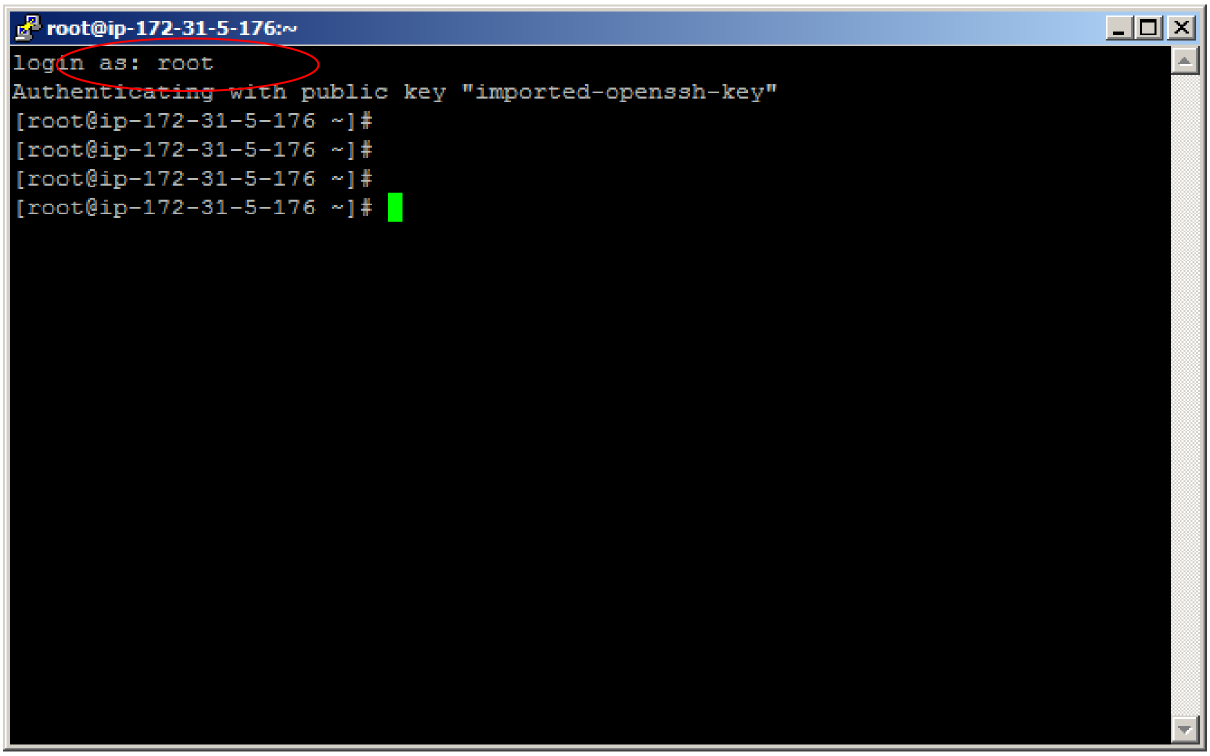
Private key file for authentication: \Amazon AWS\firsttry\HadoopExam.ppk

Clcik Open and below popup will come, just click connect.



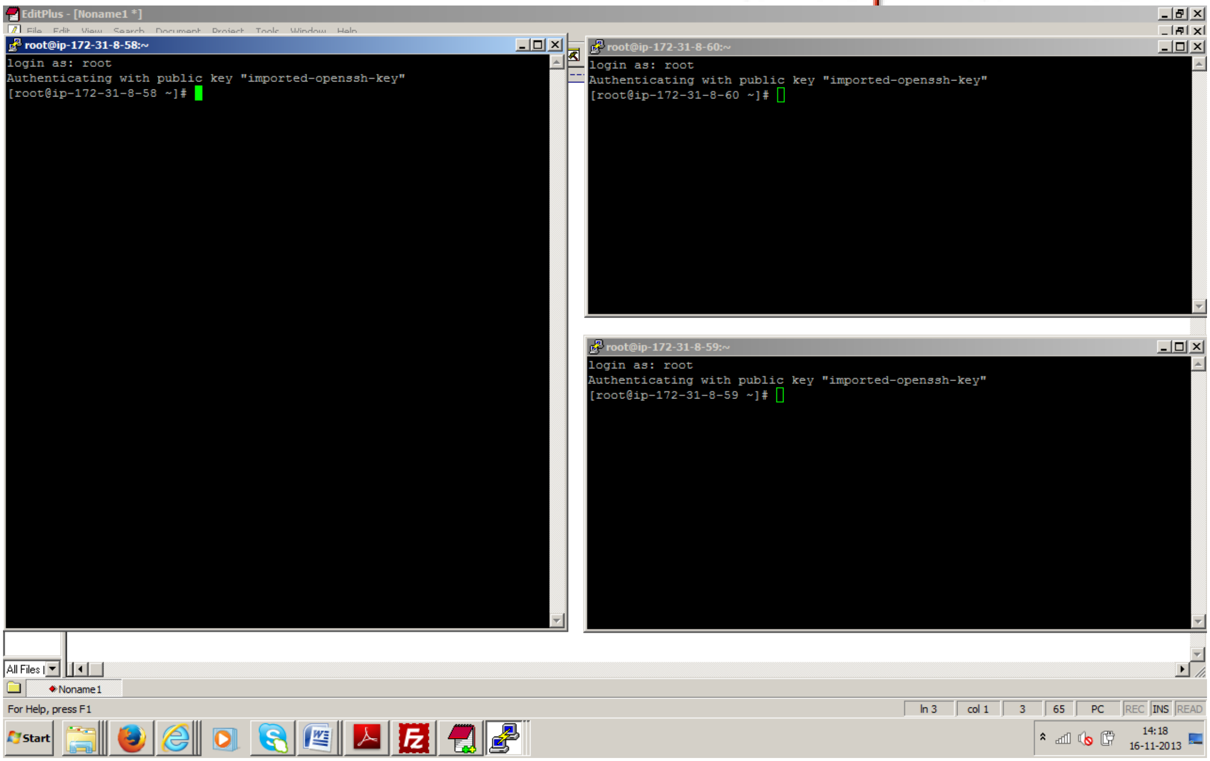
The server's host key is not cached in the registry. You have no guarantee that the server is the computer you think it is.
The server's rsa2 key fingerprint is:
ssh-rsa 2048 31:e1:cd:4f:a3:fc:01:67:b4:b3:28:f2:c0:fe:03:c2
If you trust this host, hit Yes to add the key to PuTTY's cache and carry on connecting.
If you want to carry on connecting just once, without adding the key to the cache, hit No.
If you do not trust this host, hit Cancel to abandon the connection.

Put the password as **root (Which is same for all)** and repeat the same step for remaining two nodes



```
login as: root
Authenticating with public key "imported-openssh-key"
[root@ip-172-31-5-176 ~]#
[root@ip-172-31-5-176 ~]#
[root@ip-172-31-5-176 ~]#
[root@ip-172-31-5-176 ~]#
```

18. Now we have all the three nodes connected as below. And next part we need to install Java and Hadoop on each Node.



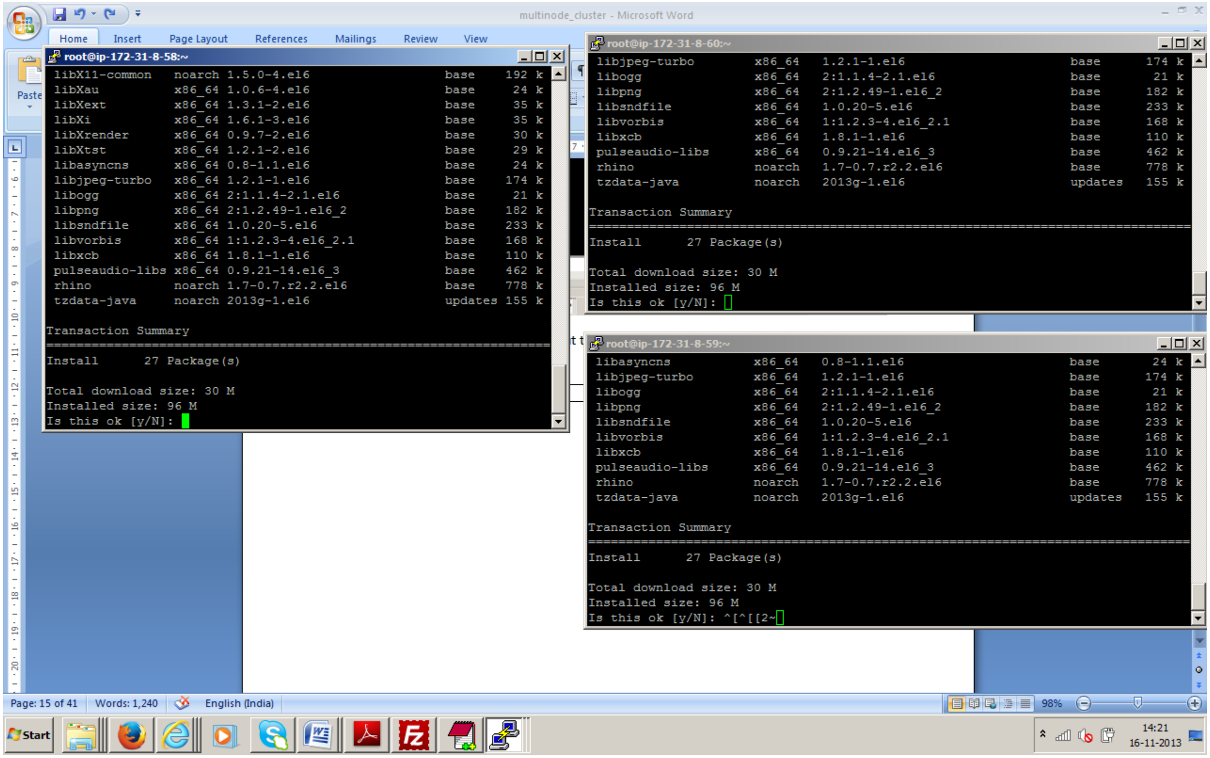
root@ip-172-31-8-58:~
```
login as: root
Authenticating with public key "imported-openssh-key"
[root@ip-172-31-8-58 ~]#
```

root@ip-172-31-8-60:~
```
login as: root
Authenticating with public key "imported-openssh-key"
[root@ip-172-31-8-60 ~]#
```

root@ip-172-31-8-59:~
```
login as: root
Authenticating with public key "imported-openssh-key"
[root@ip-172-31-8-59 ~]#
```

19. Install the Java on each Node, put the below command on each node to install openjdk.

```
yum install java-1.6.0-openjdk.x86_64
```

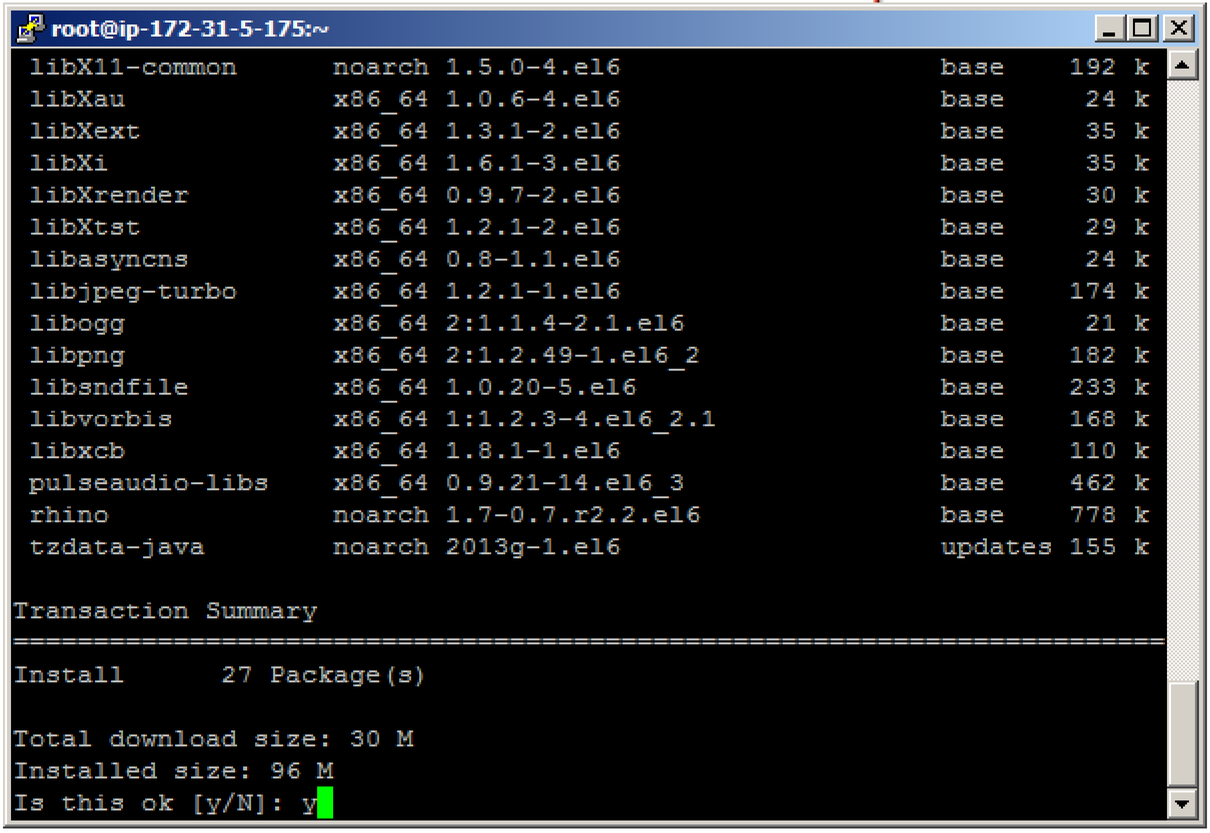


root@ip-172-31-8-58:~
```
 libX11-common   noarch 1.5.0-4.el6                          base    192 k
 libXau          x86_64 1.0.6-4.el6                          base     24 k
 libXext         x86_64 1.3.1-2.el6                          base     35 k
 libXi           x86_64 1.6.1-3.el6                          base     35 k
 libXrender      x86_64 0.9.7-2.el6                          base     30 k
 libXtst         x86_64 1.2.1-2.el6                          base     29 k
 libasyncns      x86_64 0.8-1.1.el6                          base     24 k
 libjpeg-turbo   x86_64 1.2.1-1.el6                          base    174 k
 libogg          x86_64 2:1.1.4-2.1.el6                      base     21 k
 libpng          x86_64 2:1.2.49-1.el6_2                     base    182 k
 libsndfile      x86_64 1.0.20-5.el6                         base    233 k
 libvorbis       x86_64 1:1.2.3-4.el6_2.1                    base    168 k
 libxcb          x86_64 1.8.1-1.el6                          base    110 k
 pulseaudio-libs x86_64 0.9.21-14.el6_3                      base    462 k
 rhino           noarch 1.7-0.7.r2.2.el6                     base    778 k
 tzdata-java     noarch 2013g-1.el6                          updates 155 k

Transaction Summary
====================================================================================
Install       27 Package(s)

Total download size: 30 M
Installed size: 96 M
Is this ok [y/N]:
```

root@ip-172-31-8-60:~
```
 libjpeg-turbo        x86_64    1.2.1-1.el6                          base       174 k
 libogg               x86_64    2:1.1.4-2.1.el6                      base        21 k
 libpng               x86_64    2:1.2.49-1.el6_2                     base       182 k
 libsndfile           x86_64    1.0.20-5.el6                         base       233 k
 libvorbis            x86_64    1:1.2.3-4.el6_2.1                    base       168 k
 libxcb               x86_64    1.8.1-1.el6                          base       110 k
 pulseaudio-libs      x86_64    0.9.21-14.el6_3                      base       462 k
 rhino                noarch    1.7-0.7.r2.2.el6                     base       778 k
 tzdata-java          noarch    2013g-1.el6                          updates    155 k

Transaction Summary
====================================================================================
Install       27 Package(s)

Total download size: 30 M
Installed size: 96 M
Is this ok [y/N]:
```

root@ip-172-31-8-59:~
```
 libasyncns           x86_64    0.8-1.1.el6                          base        24 k
 libjpeg-turbo        x86_64    1.2.1-1.el6                          base       174 k
 libogg               x86_64    2:1.1.4-2.1.el6                      base        21 k
 libpng               x86_64    2:1.2.49-1.el6_2                     base       182 k
 libsndfile           x86_64    1.0.20-5.el6                         base       233 k
 libvorbis            x86_64    1:1.2.3-4.el6_2.1                    base       168 k
 libxcb               x86_64    1.8.1-1.el6                          base       110 k
 pulseaudio-libs      x86_64    0.9.21-14.el6_3                      base       462 k
 rhino                noarch    1.7-0.7.r2.2.el6                     base       778 k
 tzdata-java          noarch    2013g-1.el6                          updates    155 k

Transaction Summary
====================================================================================
Install       27 Package(s)

Total download size: 30 M
Installed size: 96 M
Is this ok [y/N]: ^[^[[2~
```

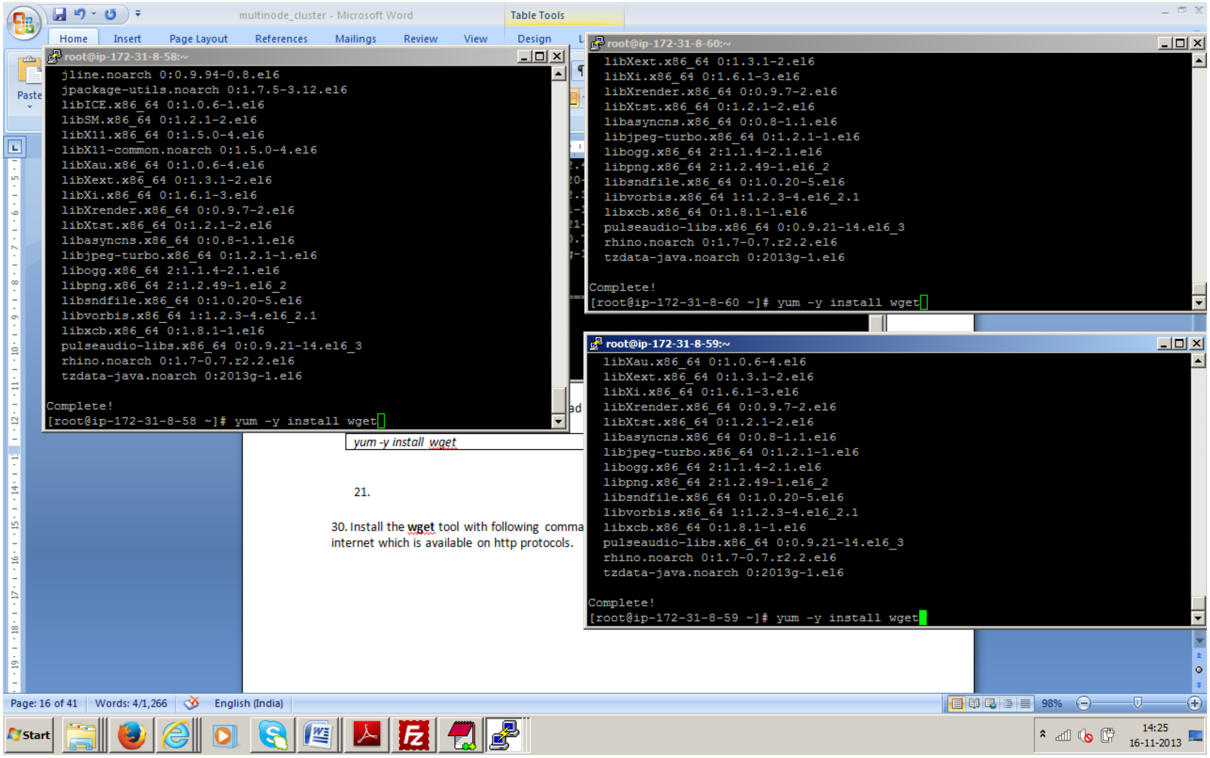
*And keep entering **y** as below in all screens.*



```
 libX11-common       noarch 1.5.0-4.el6                   base     192 k
 libXau              x86_64 1.0.6-4.el6                   base      24 k
 libXext             x86_64 1.3.1-2.el6                   base      35 k
 libXi               x86_64 1.6.1-3.el6                   base      35 k
 libXrender          x86_64 0.9.7-2.el6                   base      30 k
 libXtst             x86_64 1.2.1-2.el6                   base      29 k
 libasyncns          x86_64 0.8-1.1.el6                   base      24 k
 libjpeg-turbo       x86_64 1.2.1-1.el6                   base     174 k
 libogg              x86_64 2:1.1.4-2.1.el6               base      21 k
 libpng              x86_64 2:1.2.49-1.el6_2              base     182 k
 libsndfile          x86_64 1.0.20-5.el6                  base     233 k
 libvorbis           x86_64 1:1.2.3-4.el6_2.1             base     168 k
 libxcb              x86_64 1.8.1-1.el6                   base     110 k
 pulseaudio-libs     x86_64 0.9.21-14.el6_3               base     462 k
 rhino               noarch 1.7-0.7.r2.2.el6              base     778 k
 tzdata-java         noarch 2013g-1.el6                   updates  155 k

Transaction Summary
========================================================================
Install       27 Package(s)

Total download size: 30 M
Installed size: 96 M
Is this ok [y/N]: y
```

20. Now install another tool for download that is **wget using following command in each node.**

yum -y install wget



root@ip-172-31-8-58:~

```
  jline.noarch 0:0.9.94-0.8.el6
  jpackage-utils.noarch 0:1.7.5-3.12.el6
  libICE.x86_64 0:1.0.6-1.el6
  libSM.x86_64 0:1.2.1-2.el6
  libX11.x86_64 0:1.5.0-4.el6
  libX11-common.noarch 0:1.5.0-4.el6
  libXau.x86_64 0:1.0.6-4.el6
  libXext.x86_64 0:1.3.1-2.el6
  libXi.x86_64 0:1.6.1-3.el6
  libXrender.x86_64 0:0.9.7-2.el6
  libXtst.x86_64 0:1.2.1-2.el6
  libasyncns.x86_64 0:0.8-1.1.el6
  libjpeg-turbo.x86_64 0:1.2.1-1.el6
  libogg.x86_64 2:1.1.4-2.1.el6
  libpng.x86_64 2:1.2.49-1.el6_2
  libsndfile.x86_64 0:1.0.20-5.el6
  libvorbis.x86_64 1:1.2.3-4.el6_2.1
  libxcb.x86_64 0:1.8.1-1.el6
  pulseaudio-libs.x86_64 0:0.9.21-14.el6_3
  rhino.noarch 0:1.7-0.7.r2.2.el6
  tzdata-java.noarch 0:2013g-1.el6

Complete!
[root@ip-172-31-8-58 ~]# yum -y install wget
```

root@ip-172-31-8-60:~

```
  libXext.x86_64 0:1.3.1-2.el6
  libXi.x86_64 0:1.6.1-3.el6
  libXrender.x86_64 0:0.9.7-2.el6
  libXtst.x86_64 0:1.2.1-2.el6
  libasyncns.x86_64 0:0.8-1.1.el6
  libjpeg-turbo.x86_64 0:1.2.1-1.el6
  libogg.x86_64 2:1.1.4-2.1.el6
  libpng.x86_64 2:1.2.49-1.el6_2
  libsndfile.x86_64 0:1.0.20-5.el6
  libvorbis.x86_64 1:1.2.3-4.el6_2.1
  libxcb.x86_64 0:1.8.1-1.el6
  pulseaudio-libs.x86_64 0:0.9.21-14.el6_3
  rhino.noarch 0:1.7-0.7.r2.2.el6
  tzdata-java.noarch 0:2013g-1.el6

Complete!
[root@ip-172-31-8-60 ~]# yum -y install wget
```

root@ip-172-31-8-59:~

```
  libXau.x86_64 0:1.0.6-4.el6
  libXext.x86_64 0:1.3.1-2.el6
  libXi.x86_64 0:1.6.1-3.el6
  libXrender.x86_64 0:0.9.7-2.el6
  libXtst.x86_64 0:1.2.1-2.el6
  libasyncns.x86_64 0:0.8-1.1.el6
  libjpeg-turbo.x86_64 0:1.2.1-1.el6
  libogg.x86_64 2:1.1.4-2.1.el6
  libpng.x86_64 2:1.2.49-1.el6_2
  libsndfile.x86_64 0:1.0.20-5.el6
  libvorbis.x86_64 1:1.2.3-4.el6_2.1
  libxcb.x86_64 0:1.8.1-1.el6
  pulseaudio-libs.x86_64 0:0.9.21-14.el6_3
  rhino.noarch 0:1.7-0.7.r2.2.el6
  tzdata-java.noarch 0:2013g-1.el6

Complete!
[root@ip-172-31-8-59 ~]# yum -y install wget
```

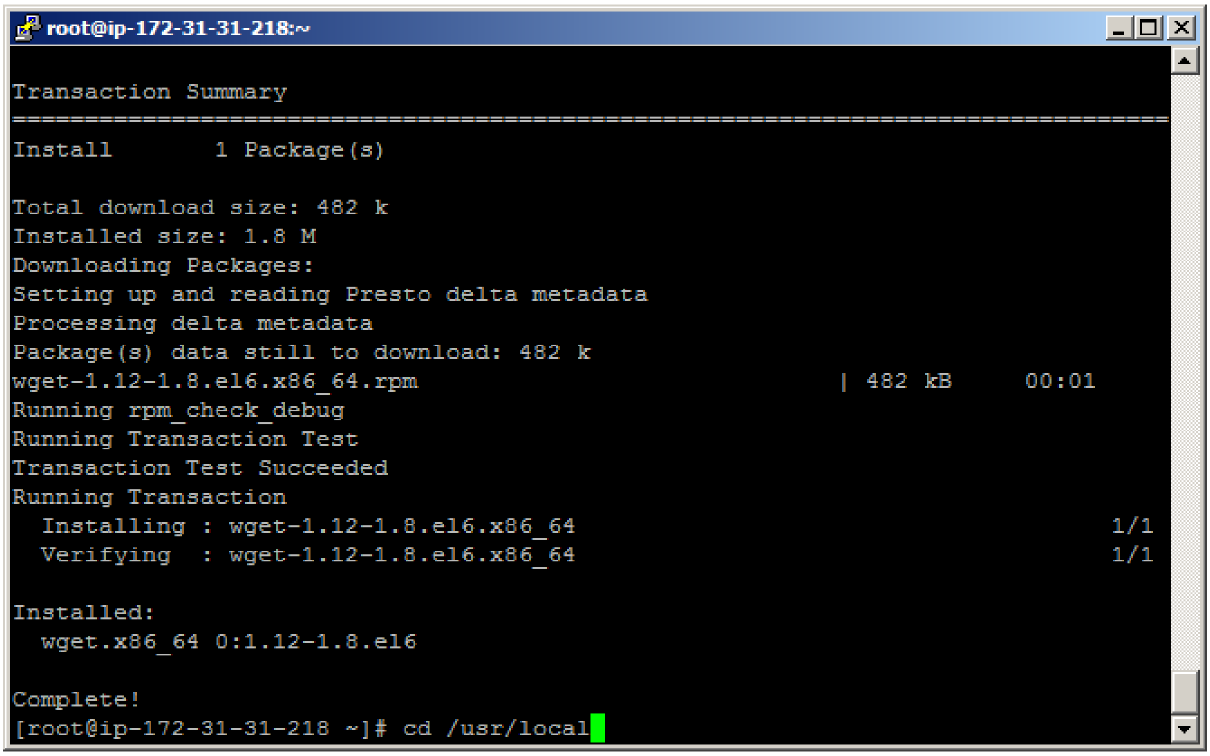
yum -y install wget

21.

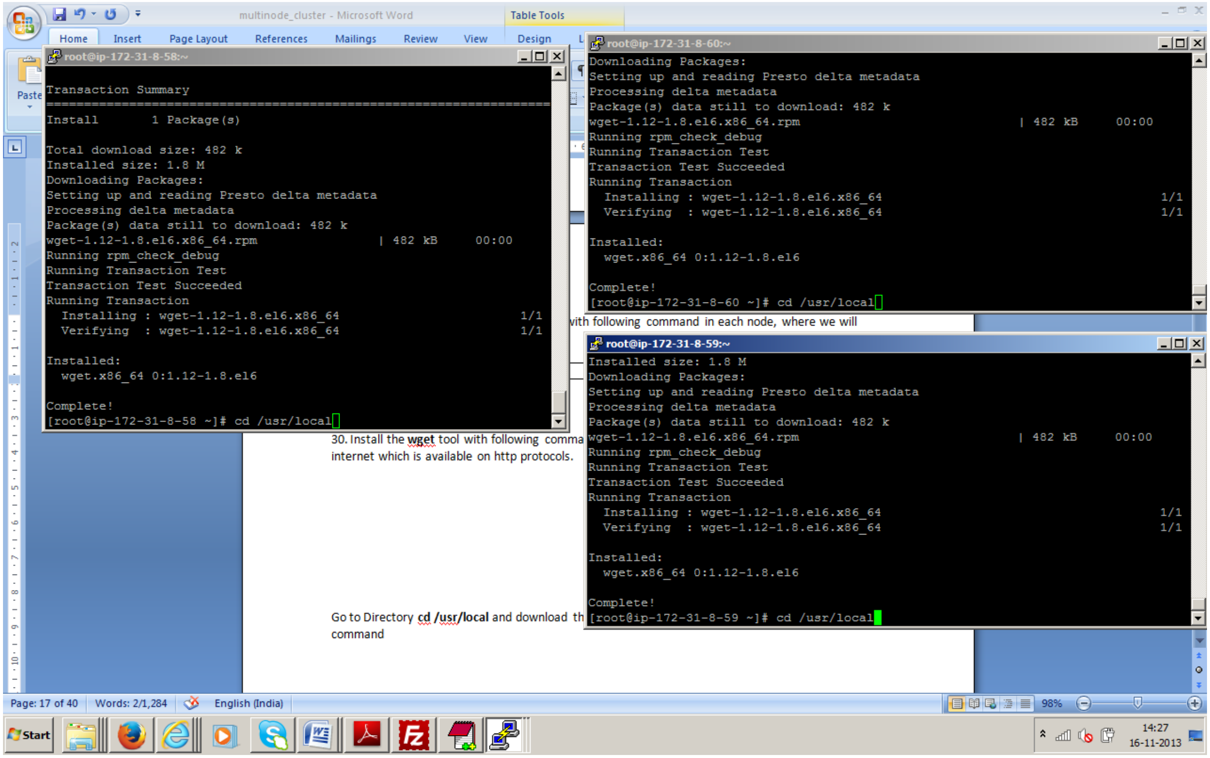
30. Install the wget tool with following comma
internet which is available on http protocols.

21. Now move to directory /usr/local/ with following command in each node, where we will install the Hadoop.

```
cd /usr/local
```



```
Transaction Summary
================================================================================
Install        1 Package(s)

Total download size: 482 k
Installed size: 1.8 M
Downloading Packages:
Setting up and reading Presto delta metadata
Processing delta metadata
Package(s) data still to download: 482 k
wget-1.12-1.8.el6.x86_64.rpm                           | 482 kB     00:01
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : wget-1.12-1.8.el6.x86_64                                     1/1
  Verifying  : wget-1.12-1.8.el6.x86_64                                     1/1

Installed:
  wget.x86_64 0:1.12-1.8.el6

Complete!
[root@ip-172-31-31-218 ~]# cd /usr/local
```



root@ip-172-31-8-58:~

```
Transaction Summary
================================================================================
Install        1 Package(s)

Total download size: 482 k
Installed size: 1.8 M
Downloading Packages:
Setting up and reading Presto delta metadata
Processing delta metadata
Package(s) data still to download: 482 k
wget-1.12-1.8.el6.x86_64.rpm                           | 482 kB     00:00
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : wget-1.12-1.8.el6.x86_64                                     1/1
  Verifying  : wget-1.12-1.8.el6.x86_64                                     1/1

Installed:
  wget.x86_64 0:1.12-1.8.el6

Complete!
[root@ip-172-31-8-58 ~]# cd /usr/local
```

root@ip-172-31-8-60:~

```
Downloading Packages:
Setting up and reading Presto delta metadata
Processing delta metadata
Package(s) data still to download: 482 k
wget-1.12-1.8.el6.x86_64.rpm                           | 482 kB     00:00
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : wget-1.12-1.8.el6.x86_64                                     1/1
  Verifying  : wget-1.12-1.8.el6.x86_64                                     1/1

Installed:
  wget.x86_64 0:1.12-1.8.el6

Complete!
[root@ip-172-31-8-60 ~]# cd /usr/local
```

ith following command in each node, where we will

root@ip-172-31-8-59:~

```
Installed size: 1.8 M
Downloading Packages:
Setting up and reading Presto delta metadata
Processing delta metadata
Package(s) data still to download: 482 k
wget-1.12-1.8.el6.x86_64.rpm                           | 482 kB     00:00
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : wget-1.12-1.8.el6.x86_64                                     1/1
  Verifying  : wget-1.12-1.8.el6.x86_64                                     1/1

Installed:
  wget.x86_64 0:1.12-1.8.el6

Complete!
[root@ip-172-31-8-59 ~]# cd /usr/local
```

30. Install the **wget** tool with following comma
internet which is available on http protocols.

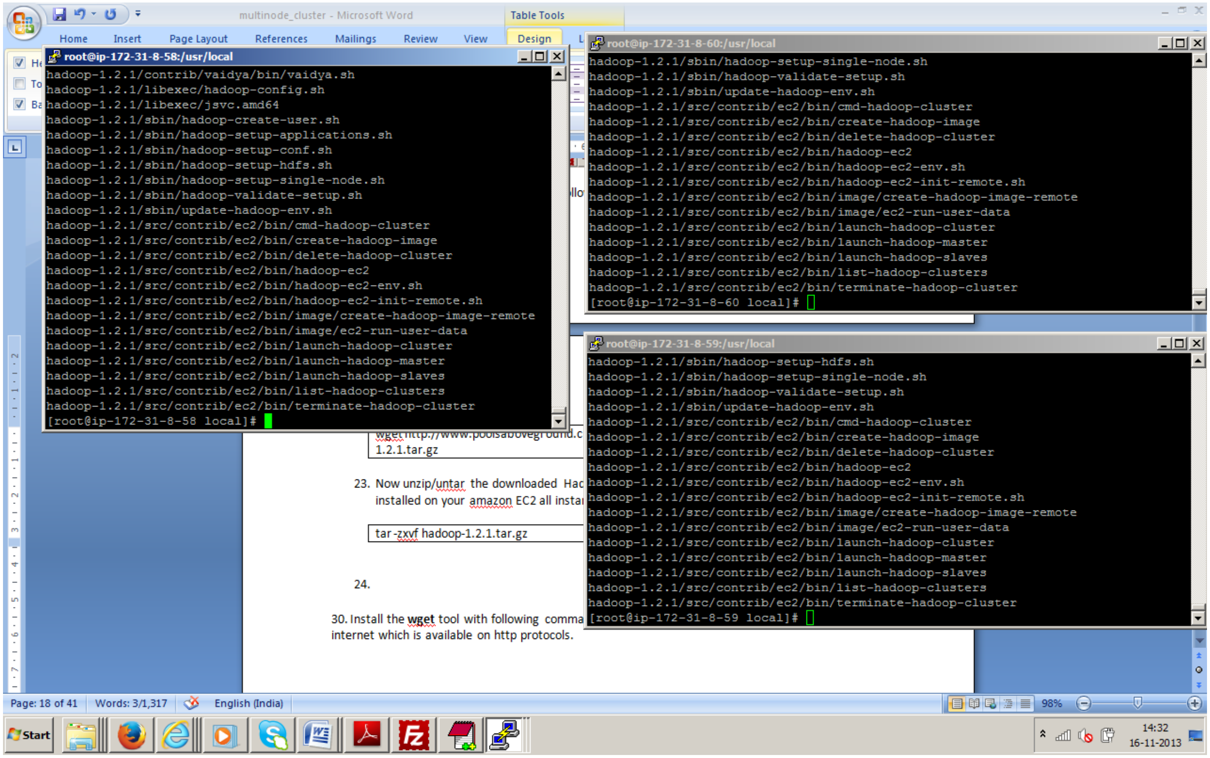
Go to Directory **cd /usr/local** and download th
command

22. Now download the Hadoop using following command.

```
wget http://www.poolsaboveground.com/apache/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz
```

23. Now unzip/untar the downloaded Hadoop framework with command. And now Hadoop is installed on your amazon EC2 all instance.

```
tar -zxvf hadoop-1.2.1.tar.gz
```



```
root@ip-172-31-8-58:/usr/local
hadoop-1.2.1/contrib/vaidya/bin/vaidya.sh
hadoop-1.2.1/libexec/hadoop-config.sh
hadoop-1.2.1/libexec/jsvc.amd64
hadoop-1.2.1/sbin/hadoop-create-user.sh
hadoop-1.2.1/sbin/hadoop-setup-applications.sh
hadoop-1.2.1/sbin/hadoop-setup-conf.sh
hadoop-1.2.1/sbin/hadoop-setup-hdfs.sh
hadoop-1.2.1/sbin/hadoop-setup-single-node.sh
hadoop-1.2.1/sbin/hadoop-validate-setup.sh
hadoop-1.2.1/sbin/update-hadoop-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/cmd-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/create-hadoop-image
hadoop-1.2.1/src/contrib/ec2/bin/delete-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-init-remote.sh
hadoop-1.2.1/src/contrib/ec2/bin/image/create-hadoop-image-remote
hadoop-1.2.1/src/contrib/ec2/bin/image/ec2-run-user-data
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-master
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-slaves
hadoop-1.2.1/src/contrib/ec2/bin/list-hadoop-clusters
hadoop-1.2.1/src/contrib/ec2/bin/terminate-hadoop-cluster
[root@ip-172-31-8-58 local]#
```

```
root@ip-172-31-8-60:/usr/local
hadoop-1.2.1/sbin/hadoop-setup-single-node.sh
hadoop-1.2.1/sbin/hadoop-validate-setup.sh
hadoop-1.2.1/sbin/update-hadoop-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/cmd-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/create-hadoop-image
hadoop-1.2.1/src/contrib/ec2/bin/delete-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-init-remote.sh
hadoop-1.2.1/src/contrib/ec2/bin/image/create-hadoop-image-remote
hadoop-1.2.1/src/contrib/ec2/bin/image/ec2-run-user-data
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-master
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-slaves
hadoop-1.2.1/src/contrib/ec2/bin/list-hadoop-clusters
hadoop-1.2.1/src/contrib/ec2/bin/terminate-hadoop-cluster
[root@ip-172-31-8-60 local]#
```

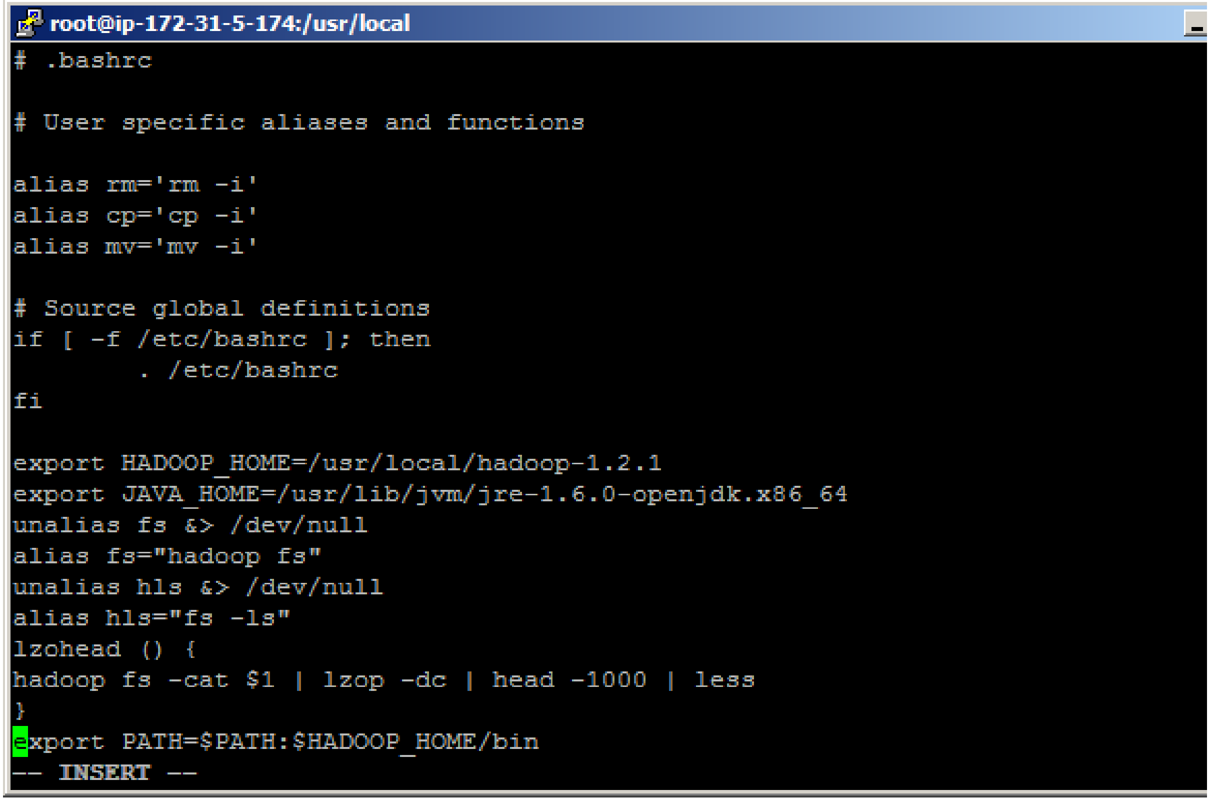
```
root@ip-172-31-8-59:/usr/local
hadoop-1.2.1/sbin/hadoop-setup-hdfs.sh
hadoop-1.2.1/sbin/hadoop-setup-single-node.sh
hadoop-1.2.1/sbin/hadoop-validate-setup.sh
hadoop-1.2.1/sbin/update-hadoop-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/cmd-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/create-hadoop-image
hadoop-1.2.1/src/contrib/ec2/bin/delete-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-env.sh
hadoop-1.2.1/src/contrib/ec2/bin/hadoop-ec2-init-remote.sh
hadoop-1.2.1/src/contrib/ec2/bin/image/create-hadoop-image-remote
hadoop-1.2.1/src/contrib/ec2/bin/image/ec2-run-user-data
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-cluster
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-master
hadoop-1.2.1/src/contrib/ec2/bin/launch-hadoop-slaves
hadoop-1.2.1/src/contrib/ec2/bin/list-hadoop-clusters
hadoop-1.2.1/src/contrib/ec2/bin/terminate-hadoop-cluster
[root@ip-172-31-8-59 local]#
```

24. Now set the JAVA_HOME and HADOOP_HOME in the root/.bashrc file, by copying the following content on each node. As you know we have already installed Java and Hadoop in previous steps. Make sure you put proper path for java and Hadoop where it is installed. And save it by pressing, the way you save file in linux. (Do same thing on each node)

esc:wq

```
vi /root/.bashrc
```

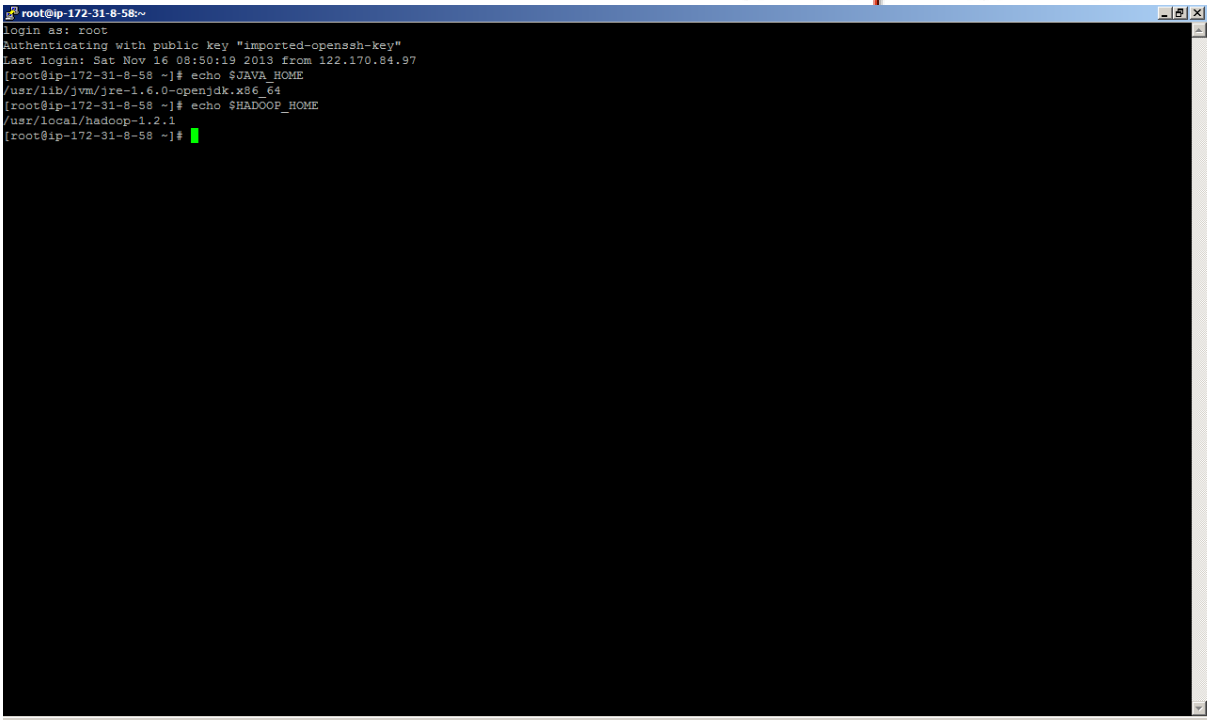
```
export HADOOP_HOME=/usr/local/hadoop-1.2.1
export JAVA_HOME=/usr/lib/jvm/jre-1.6.0-openjdk.x86_64
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"
lzohead () {
hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
export PATH=$PATH:$HADOOP_HOME/bin
```



```
# .bashrc

# User specific aliases and functions

alias rm='rm -i'
alias cp='cp -i'
alias mv='mv -i'

# Source global definitions
if [ -f /etc/bashrc ]; then
        . /etc/bashrc
fi

export HADOOP_HOME=/usr/local/hadoop-1.2.1
export JAVA_HOME=/usr/lib/jvm/jre-1.6.0-openjdk.x86_64
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"
lzohead () {
hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
export PATH=$PATH:$HADOOP_HOME/bin
-- INSERT --
```

25. Now restart putty shell to take effect this configuration and after restart JAVA_HOME and HADOOP_HOME should be available. And by typing following command you can make sure whether JAVA_HOME and HADOOP_HOME are pointing the installed location or not.

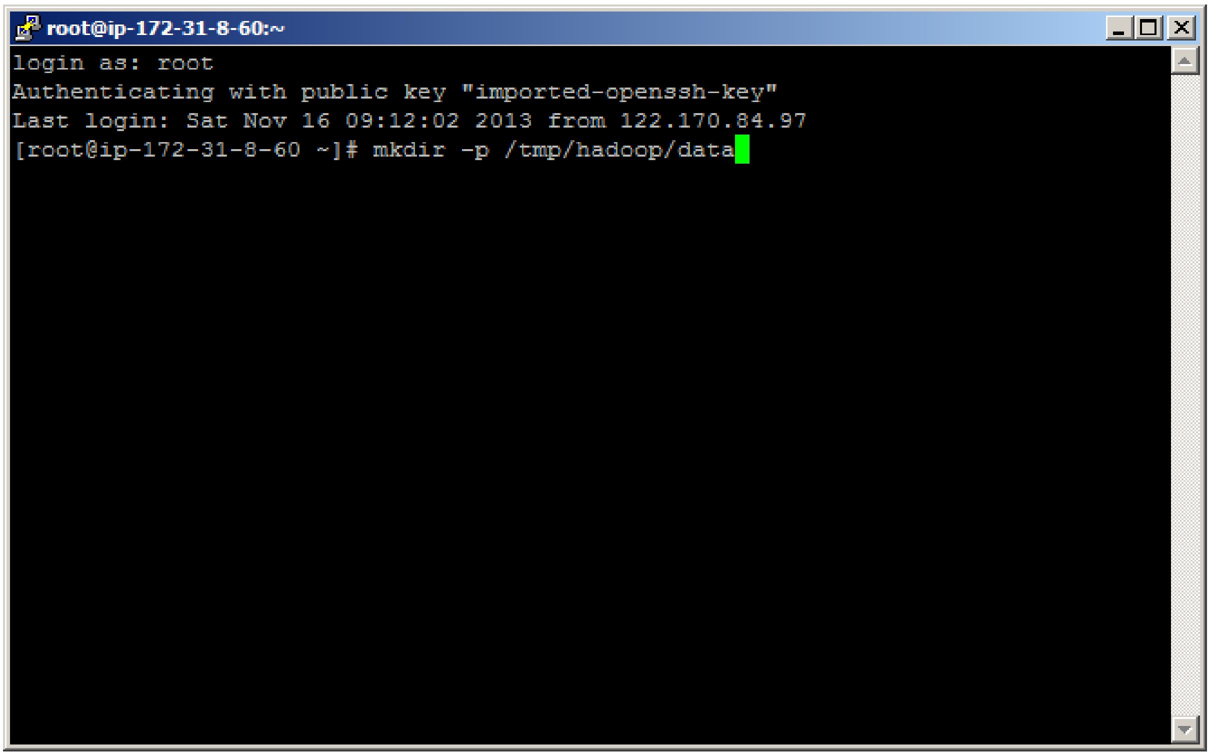
```
echo $JAVA_HOME
echo $HADOOP_HOME
```



```
login as: root
Authenticating with public key "imported-openssh-key"
Last login: Sat Nov 16 08:50:19 2013 from 122.170.84.97
[root@ip-172-31-8-58 ~]# echo $JAVA_HOME
/usr/lib/jvm/jre-1.6.0-openjdk.x86_64
[root@ip-172-31-8-58 ~]# echo $HADOOP_HOME
/usr/local/hadoop-1.2.1
[root@ip-172-31-8-58 ~]#
```

26. Create temp directory for Hadoop Data storage. So here your all data will be stored, which you will be storing in hdfs file sytem on each node.

```
mkdir -p /tmp/hadoop/data
```

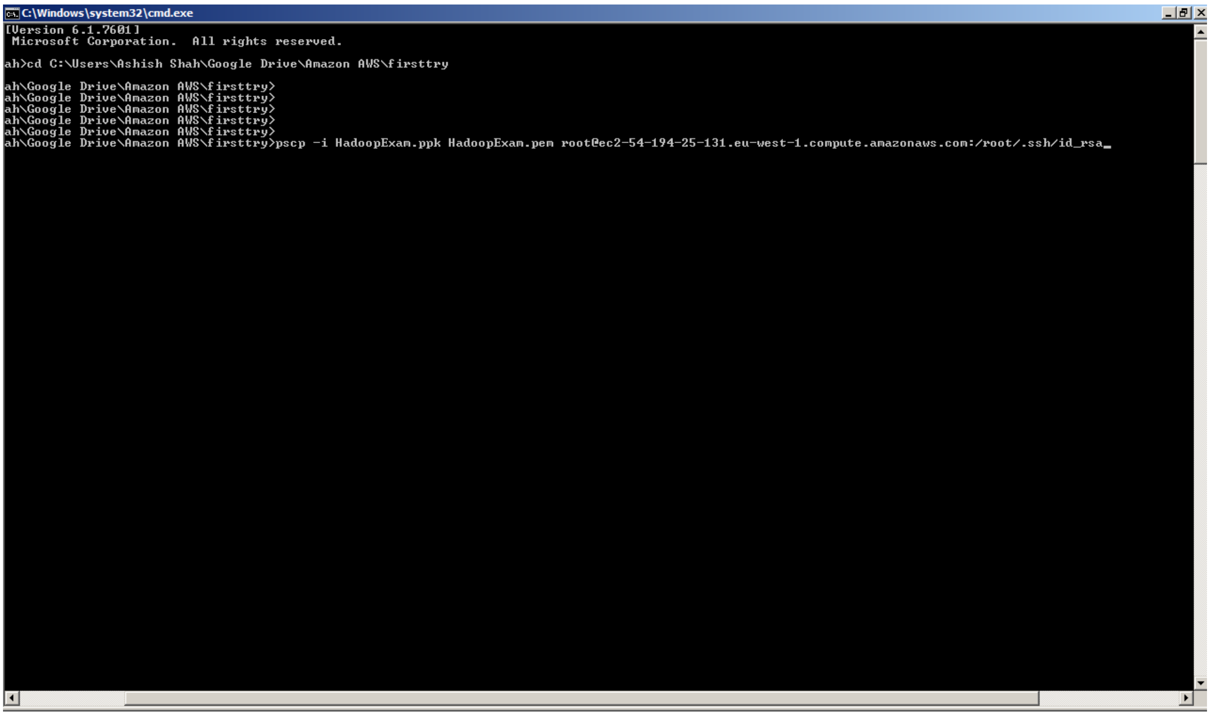


```
login as: root
Authenticating with public key "imported-openssh-key"
Last login: Sat Nov 16 09:12:02 2013 from 122.170.84.97
[root@ip-172-31-8-60 ~]# mkdir -p /tmp/hadoop/data
```

27. Now we have 3 nodes running but they cannot communicate with each other and to work our cluster we want all nodes should be able to communicate with each other. So we need to setup password less ssh communication.
28. Now open the command prompt and using pscp tool we will copy the Private key (.ppk and .pem files) from windows machine to ***master node only.***

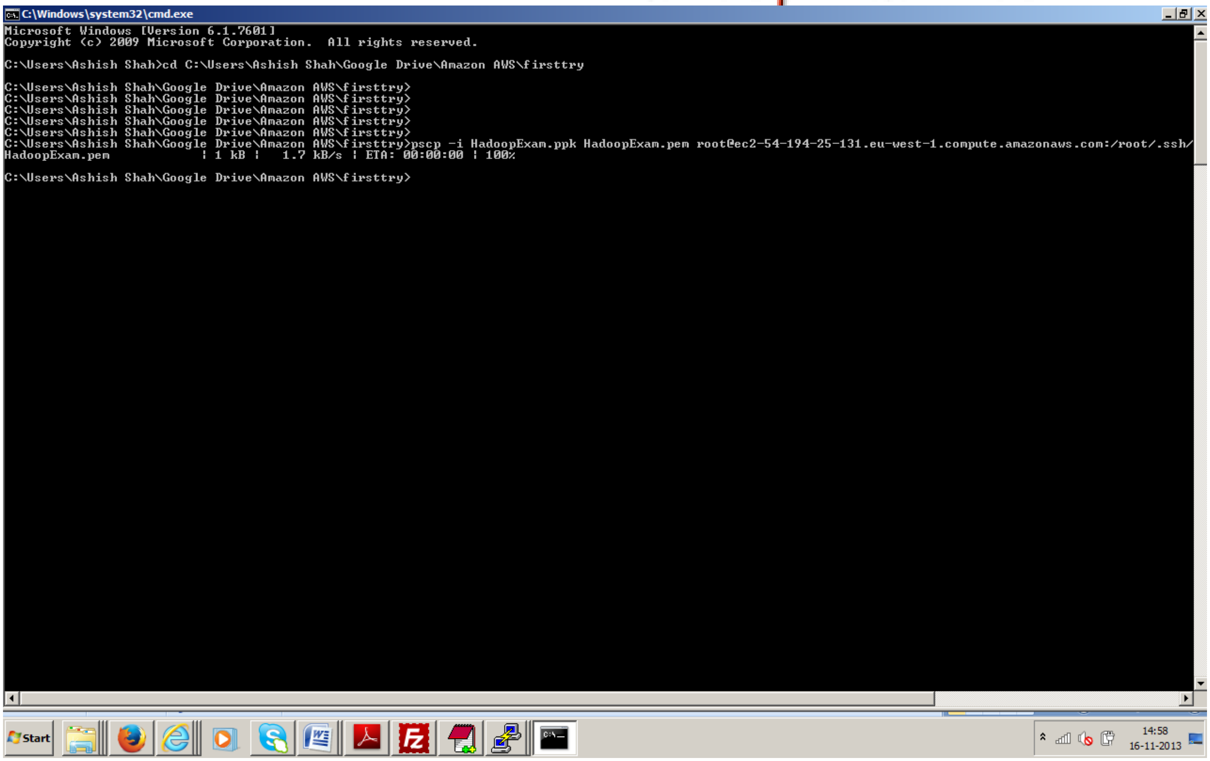
- Open Windows Start
- Run: cmd (command line utility)
- Navigate to where the HadoopExam.ppk and HadoopExam.pem are located, for us it was c:\ (root)
- Make sure the pscp.exe (putty scp) file is downloaded to this directory or its path is part of the windows global variable so it can be executed from any directory

```
pscp -i HadoopExam.ppk HadoopExam.pem root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/root/.ssh/id_rsa
```



C:\Windows\system32\cmd.exe

```
[Version 6.1.7601]
 Microsoft Corporation.  All rights reserved.

ah>cd C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry

ah\Google Drive\Amazon AWS\firsttry>
ah\Google Drive\Amazon AWS\firsttry>
ah\Google Drive\Amazon AWS\firsttry>
ah\Google Drive\Amazon AWS\firsttry>
ah\Google Drive\Amazon AWS\firsttry>
ah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.pem root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/root/.ssh/id_rsa_
```

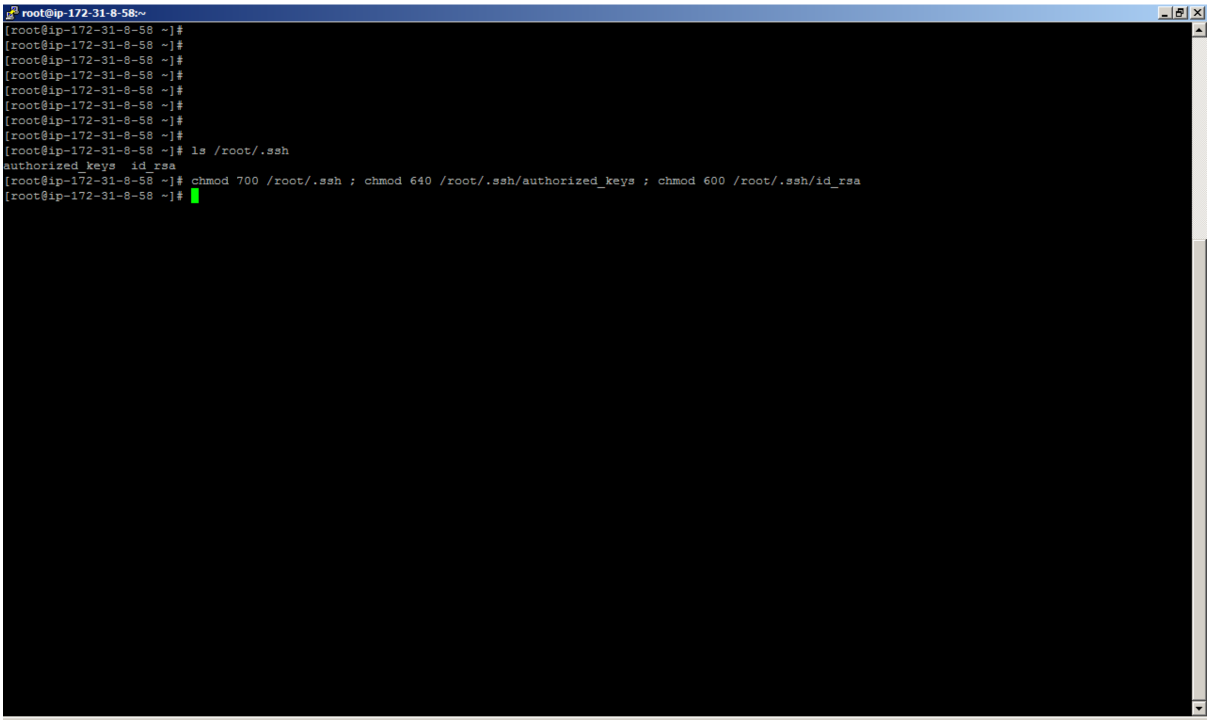


```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Ashish Shah>cd C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.pem root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/root/.ssh/
HadoopExam.pem            | 1 kB |   1.7 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
```

29. Set permissions on the master node for using the private key file

```
ls /root/.ssh
chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
```

This well Set /root/.ssh directory to owner=execute ; set public key file 'authorized keys' to owner=read+execute,group=read ; and set private key file 'id_rsa' to owner=read+execute



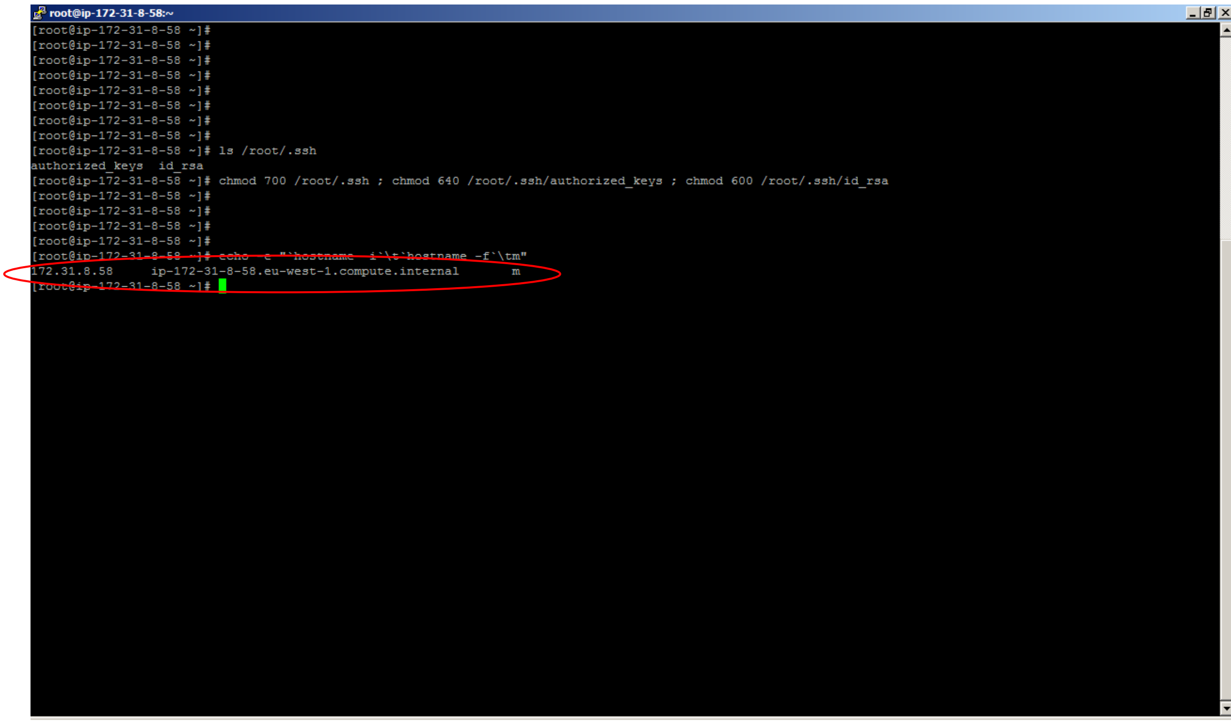
```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
```

30. Gather network information for each node as below.

- On Master Node execute following command.

```
echo -e "`hostname -i`\t`hostname -f`\tm"
```

- Print to stdout the IP address, Private address, m (host alias) using echo with special character '\t' to delimit the values with a tab character.
- save this line for the master node m to the previously created notes file as it will be used later to populate the /etc/hosts file of each instance
- You should be able to ssh from the master node **m** into the other instances (n1,n2) now that we have uploaded the .pem as id_rsa and set permissions



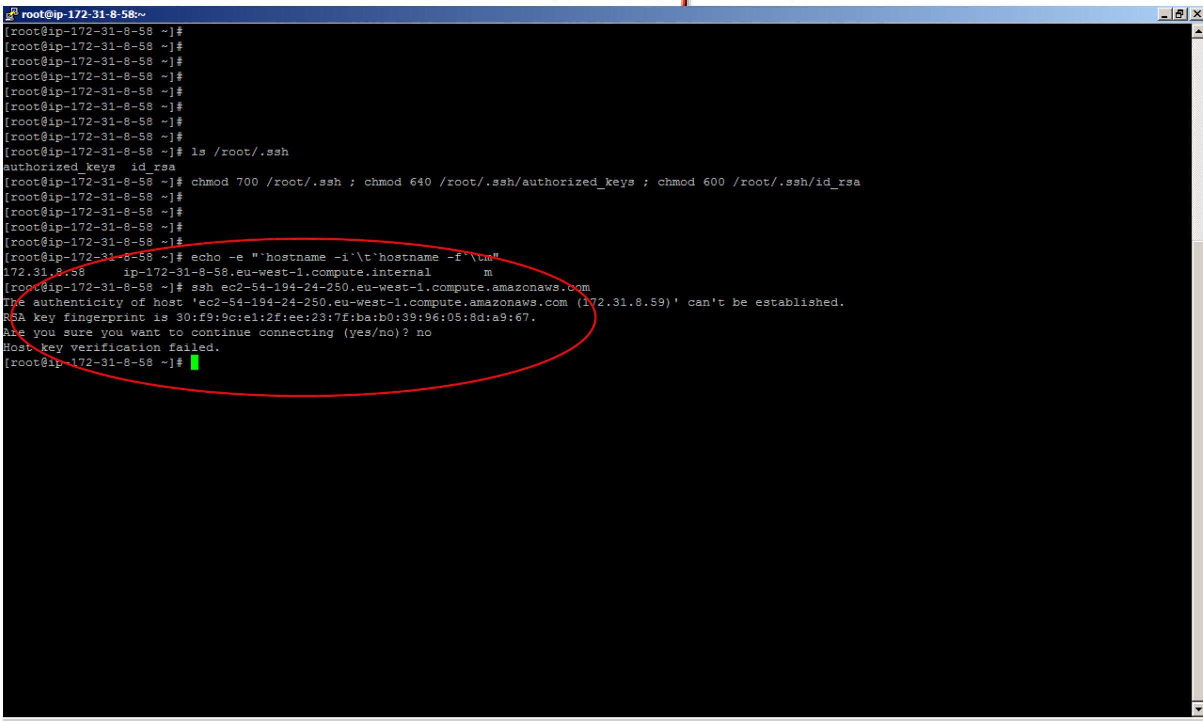
root@ip-172-31-8-58:~

```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]#
```

- From master node now we should be able to connect data node as well.

```
ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
[TYPE NO TO AUTHENTICATION REQUEST]
```

You will notice the ssh tool asking you to authenticate the host, since this is T&D lets remove this message so we can ssh automatically into all of the nodes

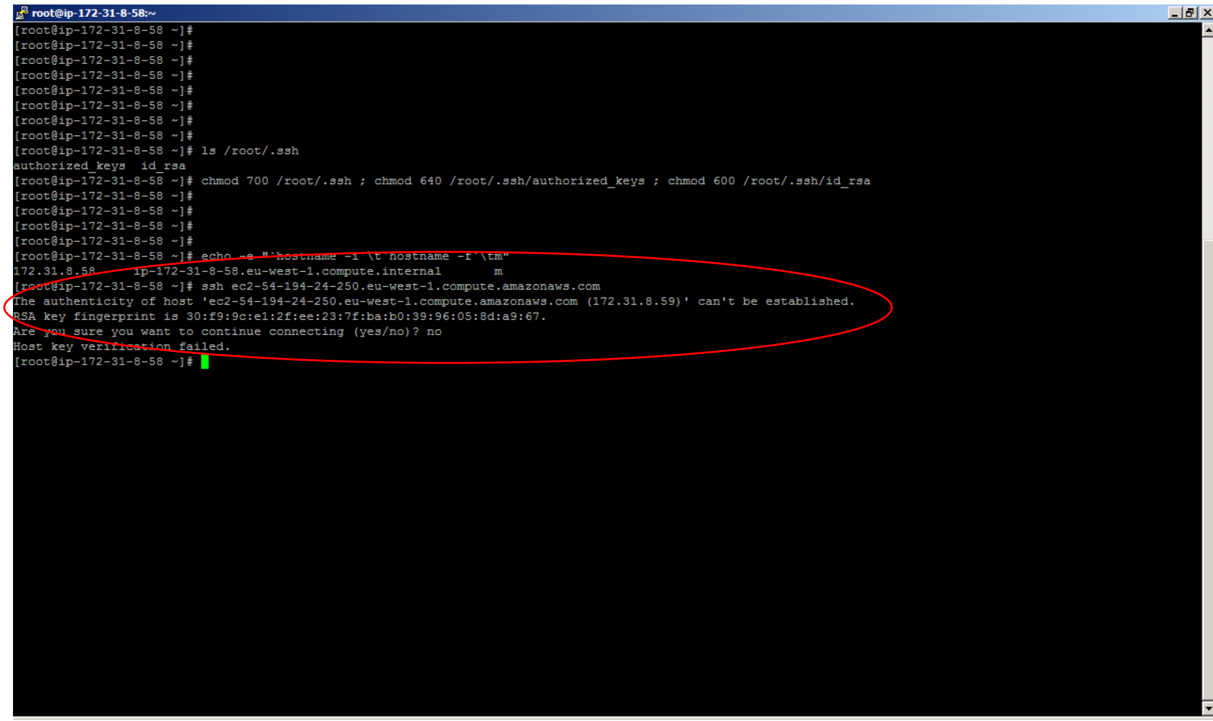


```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]#
```

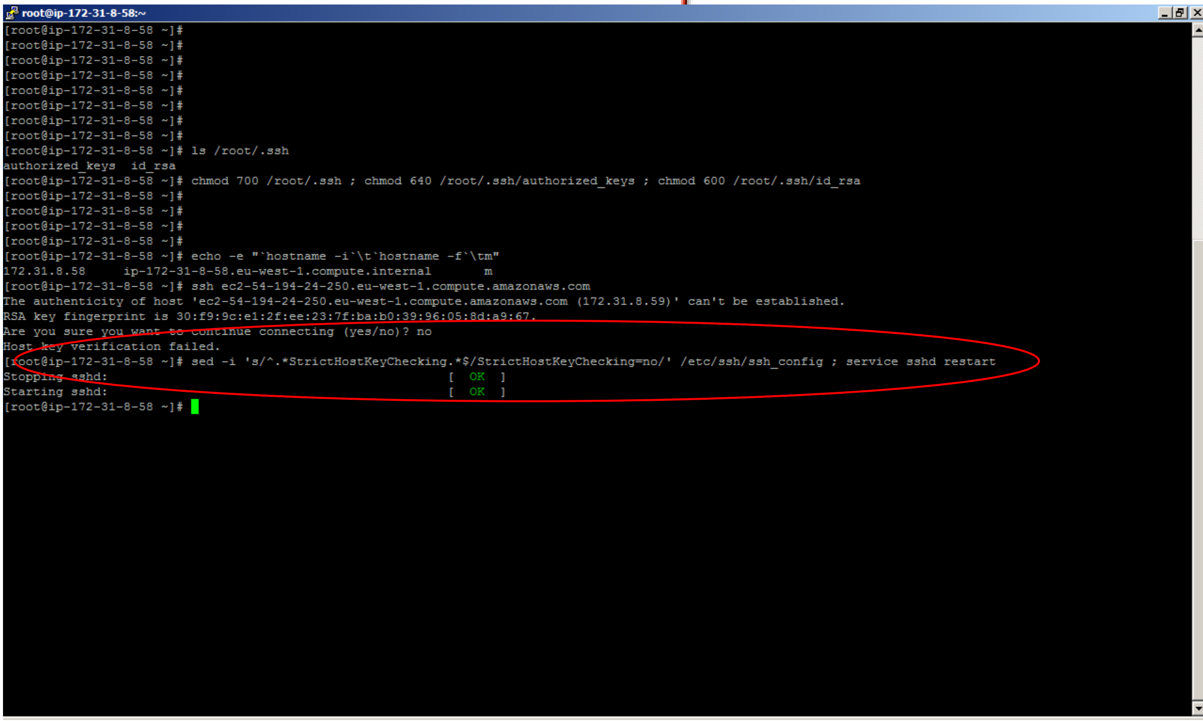
- Execute following command on master node. This will use the sed command to search and replace from the /etc/ssh/ssh_config file the first occurrence of the variable StrictHostKeyChecking and set it equal to 'no'

  The next command will restart the ssh daemon such that it can pick up this configuration file change

```
sed -i 's/^.*StrictHostKeyChecking.*$/StrictHostKeyChecking=no/' /etc/ssh/ssh_config
; service sshd restart
```



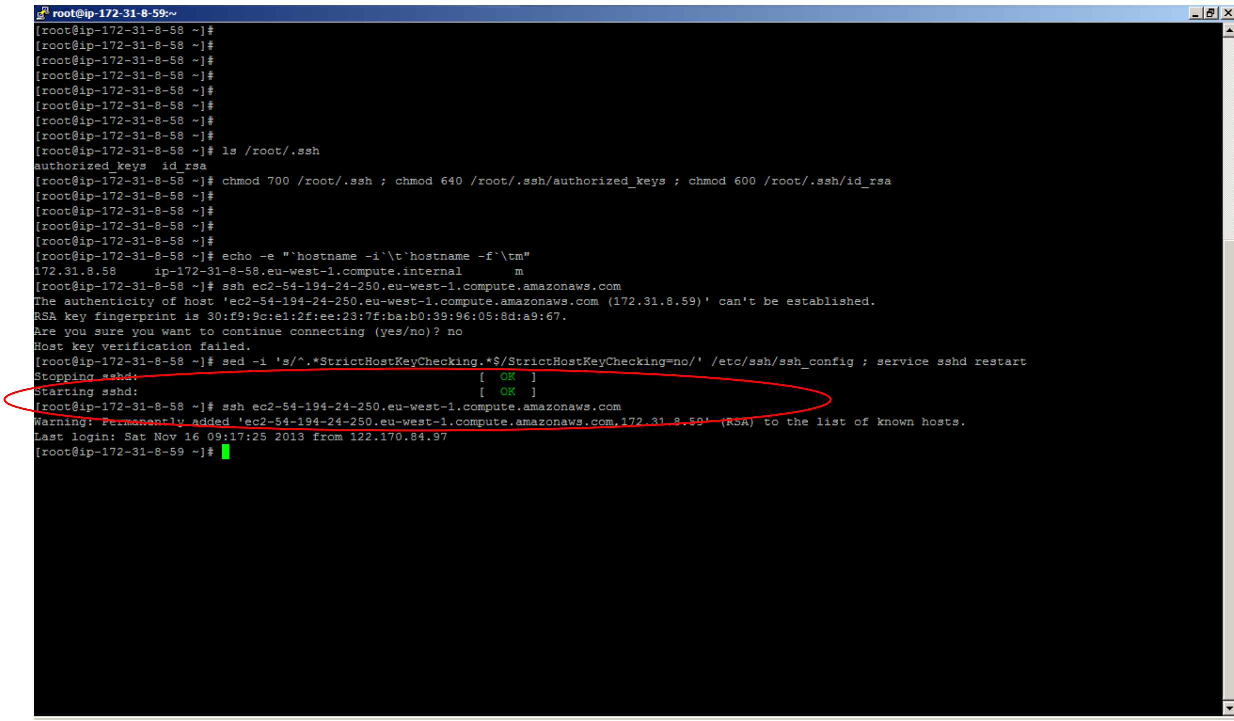
```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]#
```



```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]# sed -i 's/^.*StrictHostKeyChecking.*$/StrictHostKeyChecking=no/' /etc/ssh/ssh_config ; service sshd restart
Stopping sshd:                                             [  OK  ]
Starting sshd:                                             [  OK  ]
[root@ip-172-31-8-58 ~]#
```

- Now ssh to the datanode n1 using below command.

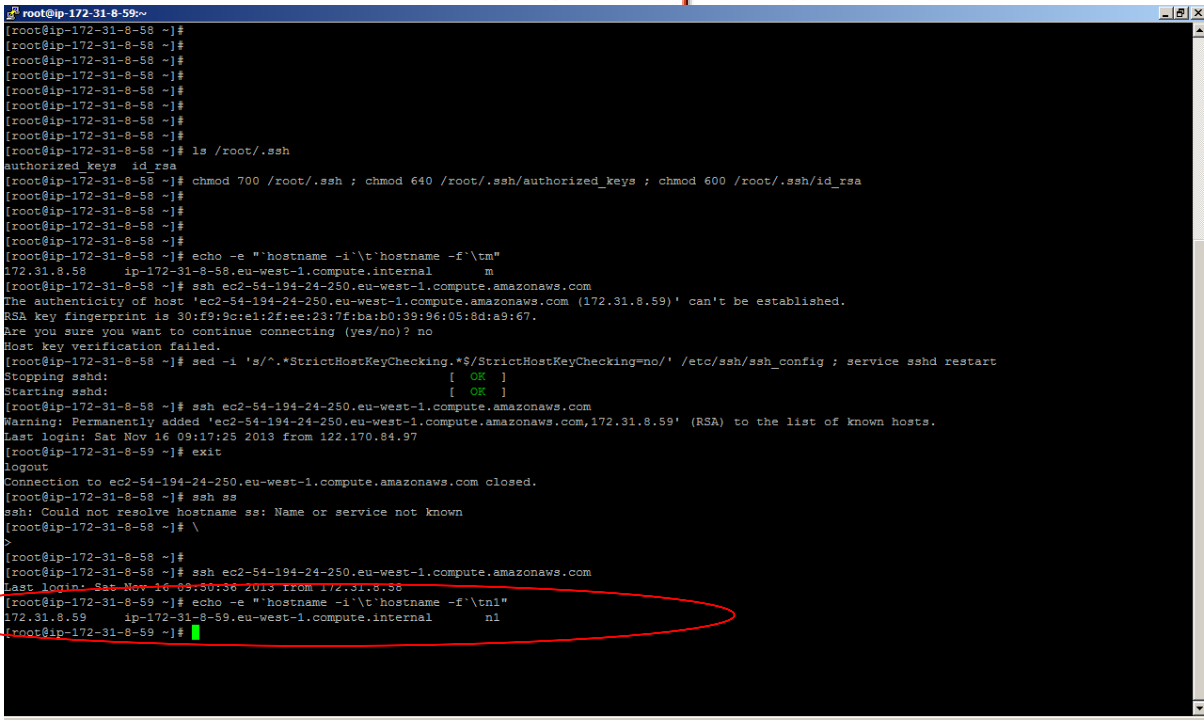
```
ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
```



```
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]# sed -i 's/^.*StrictHostKeyChecking.*$/StrictHostKeyChecking=no/' /etc/ssh/ssh_config ; service sshd restart
Stopping sshd:                                             [  OK  ]
Starting sshd:                                             [  OK  ]
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
Warning: Permanently added 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com,172.31.8.59' (RSA) to the list of known hosts.
Last login: Sat Nov 16 09:17:25 2013 from 122.170.84.97
[root@ip-172-31-8-59 ~]#
```

- Gather the n1 datanode network information from following command.

```
echo -e "`hostname -i`\t`hostname -f`\tn1"
```

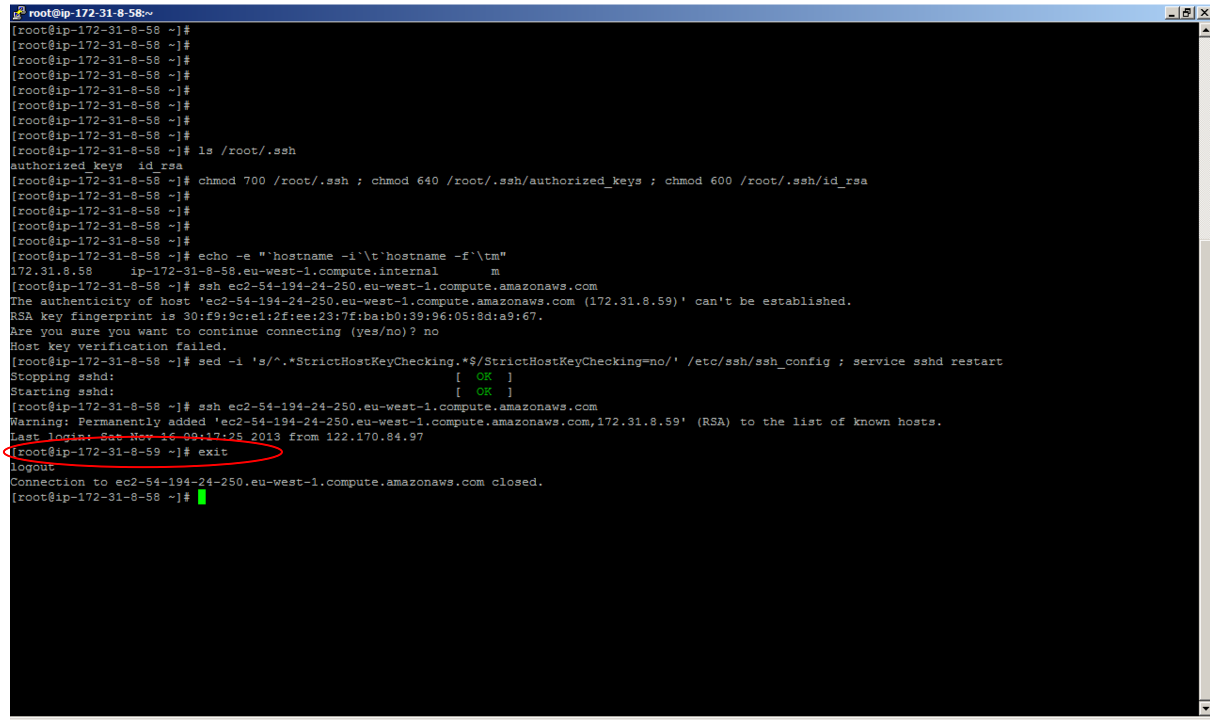


```
root@ip-172-31-8-59:~
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]# sed -i 's/^.*StrictHostKeyChecking.*$/StrictHostKeyChecking=no/' /etc/ssh/ssh_config ; service sshd restart
Stopping sshd:                                             [  OK  ]
Starting sshd:                                             [  OK  ]
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
Warning: Permanently added 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com,172.31.8.59' (RSA) to the list of known hosts.
Last login: Sat Nov 16 09:17:25 2013 from 122.170.84.97
[root@ip-172-31-8-59 ~]# exit
logout
Connection to ec2-54-194-24-250.eu-west-1.compute.amazonaws.com closed.
[root@ip-172-31-8-58 ~]# ssh ss
ssh: Could not resolve hostname ss: Name or service not known
[root@ip-172-31-8-58 ~]# \
>
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
Last login: Sat Nov 16 09:50:36 2013 from 172.31.8.58
[root@ip-172-31-8-59 ~]# echo -e "`hostname -i`\t`hostname -f`\tn1"
172.31.8.59     ip-172-31-8-59.eu-west-1.compute.internal       n1
[root@ip-172-31-8-59 ~]#
```

save this line for node1/n1 to the previously created notes file

- And now exit from datanode n1 using following command.

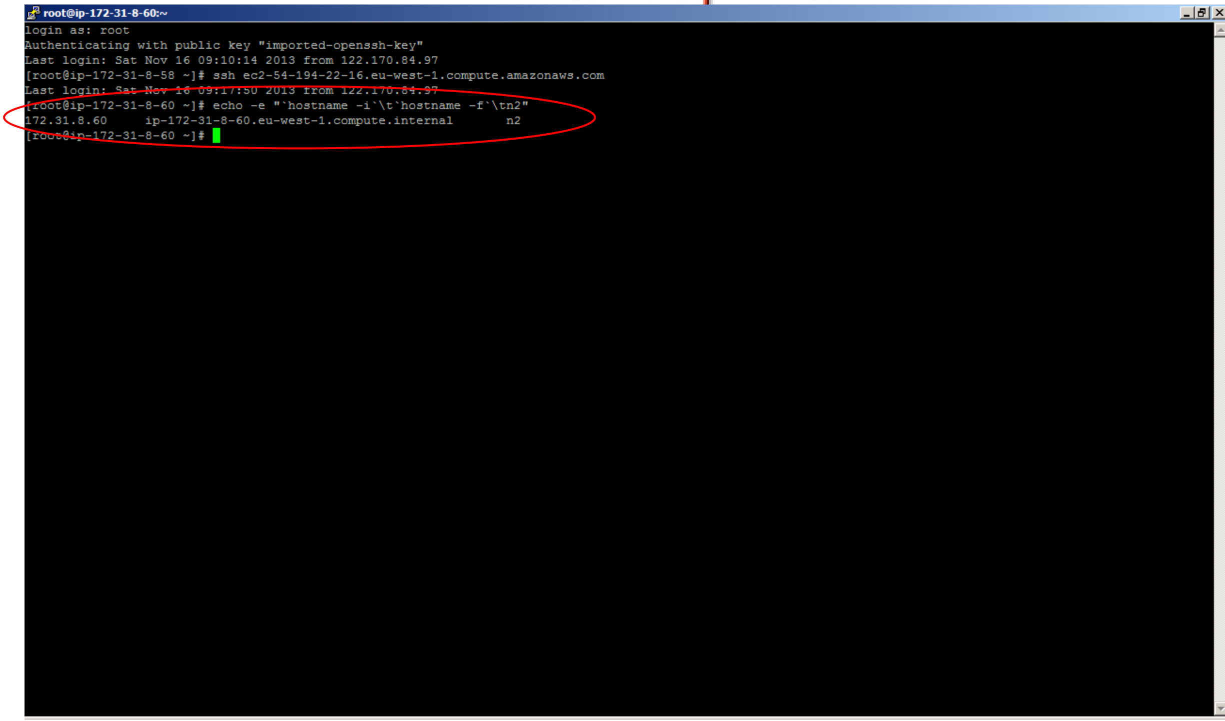
```
exit
```



```
root@ip-172-31-8-58:~
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ls /root/.ssh
authorized_keys  id_rsa
[root@ip-172-31-8-58 ~]# chmod 700 /root/.ssh ; chmod 640 /root/.ssh/authorized_keys ; chmod 600 /root/.ssh/id_rsa
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# echo -e "`hostname -i`\t`hostname -f`\tm"
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
The authenticity of host 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com (172.31.8.59)' can't be established.
RSA key fingerprint is 30:f9:9c:e1:2f:ee:23:7f:ba:b0:39:96:05:8d:a9:67.
Are you sure you want to continue connecting (yes/no)? no
Host key verification failed.
[root@ip-172-31-8-58 ~]# sed -i 's/^.*StrictHostKeyChecking.*$/StrictHostKeyChecking=no/' /etc/ssh/ssh_config ; service sshd restart
Stopping sshd:                                             [  OK  ]
Starting sshd:                                             [  OK  ]
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
Warning: Permanently added 'ec2-54-194-24-250.eu-west-1.compute.amazonaws.com,172.31.8.59' (RSA) to the list of known hosts.
Last login: Sat Nov 16 09:17:25 2013 from 122.170.84.97
[root@ip-172-31-8-59 ~]# exit
logout
Connection to ec2-54-194-24-250.eu-west-1.compute.amazonaws.com closed.
[root@ip-172-31-8-58 ~]#
```

- Now ssh to second datanode using following command and get the network information for node2/n2

```
ssh ec2-54-194-22-16.eu-west-1.compute.amazonaws.com
echo -e "`hostname -i`\t`hostname -f`\tn2"
```



```
login as: root
Authenticating with public key "imported-openssh-key"
Last login: Sat Nov 16 09:10:14 2013 from 122.170.84.97
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-22-16.eu-west-1.compute.amazonaws.com
Last login: Sat Nov 16 09:17:50 2013 from 122.170.84.97
[root@ip-172-31-8-60 ~]# echo -e "`hostname -i`\t`hostname -f`\tn2"
172.31.8.60     ip-172-31-8-60.eu-west-1.compute.internal       n2
[root@ip-172-31-8-60 ~]#
```

Now we should hane network information for all three nodes as below. (This will be different in your case)

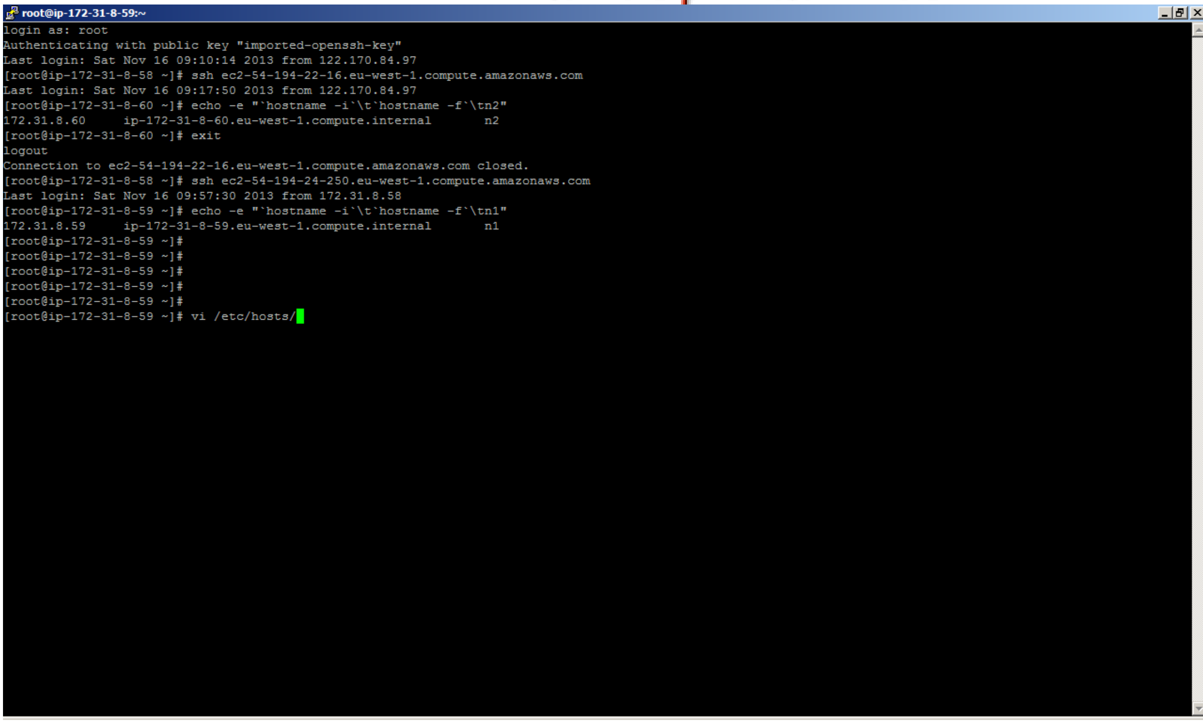
```
172.31.8.58    ip-172-31-8-58.eu-west-1.compute.internal     m
172.31.8.59    ip-172-31-8-59.eu-west-1.compute.internal     n1
172.31.8.60    ip-172-31-8-60.eu-west-1.compute.internal     n2
```

31. Now copy the above information in the etc/hosts file of each node.

On Master node

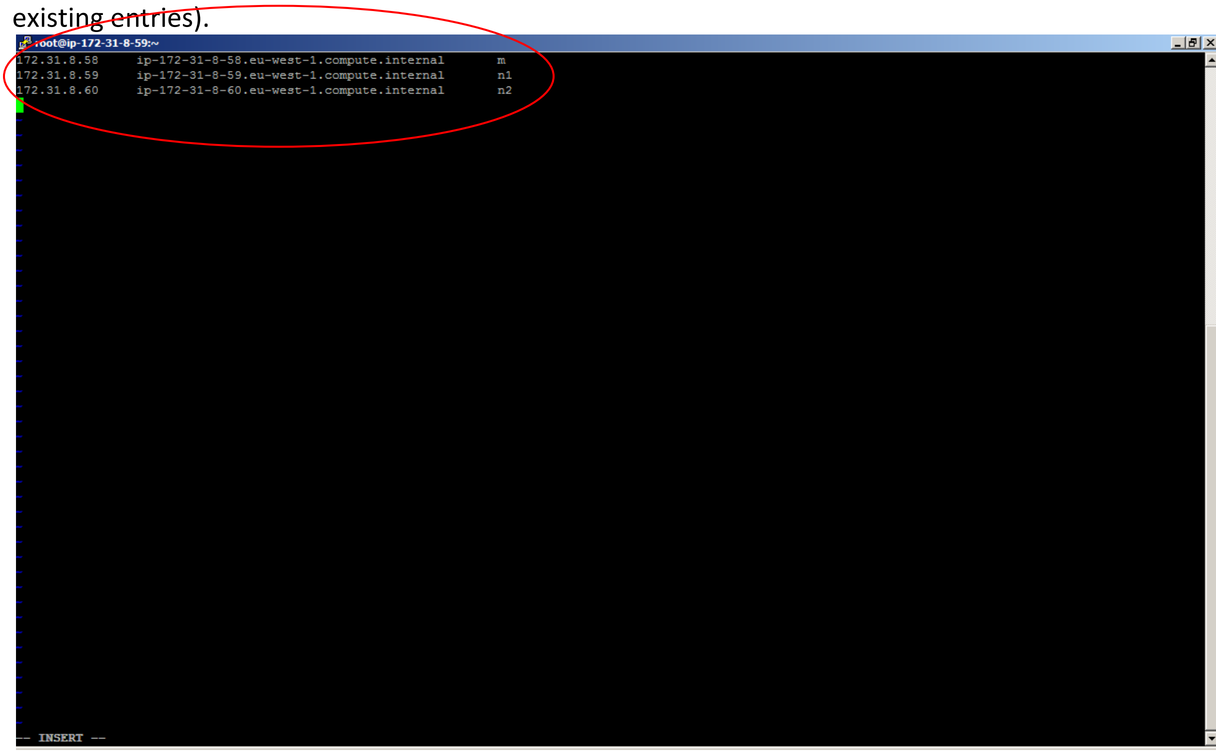
- Using the following command edit the hosts file.

```
vi /etc/hosts
```



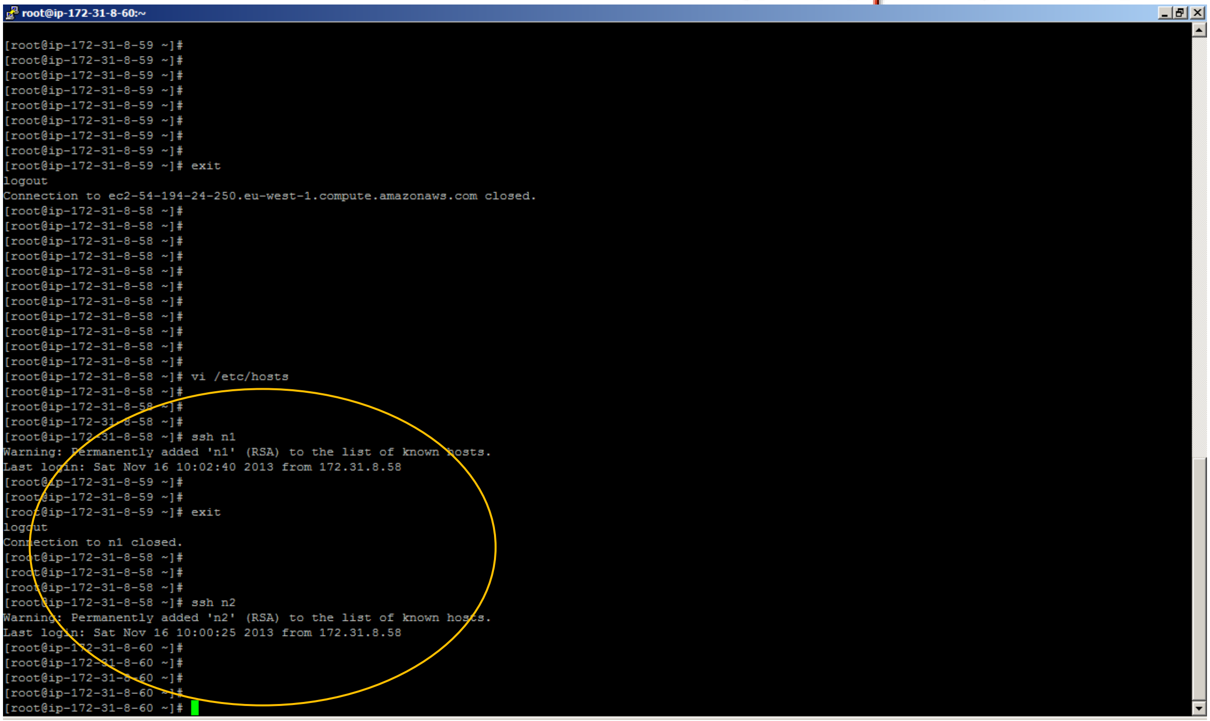
```
login as: root
Authenticating with public key "imported-openssh-key"
Last login: Sat Nov 16 09:10:14 2013 from 122.170.84.97
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-22-16.eu-west-1.compute.amazonaws.com
Last login: Sat Nov 16 09:17:50 2013 from 122.170.84.97
[root@ip-172-31-8-60 ~]# echo -e "`hostname -i`\t`hostname -f`\tn2"
172.31.8.60     ip-172-31-8-60.eu-west-1.compute.internal       n2
[root@ip-172-31-8-60 ~]# exit
logout
Connection to ec2-54-194-22-16.eu-west-1.compute.amazonaws.com closed.
[root@ip-172-31-8-58 ~]# ssh ec2-54-194-24-250.eu-west-1.compute.amazonaws.com
Last login: Sat Nov 16 09:57:30 2013 from 172.31.8.58
[root@ip-172-31-8-59 ~]# echo -e "`hostname -i`\t`hostname -f`\tn1"
172.31.8.59     ip-172-31-8-59.eu-west-1.compute.internal       n1
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]# vi /etc/hosts/
```

- And add the network information of all the nodes in this file as below(delete all the existing entries).



```
172.31.8.58     ip-172-31-8-58.eu-west-1.compute.internal       m
172.31.8.59     ip-172-31-8-59.eu-west-1.compute.internal       n1
172.31.8.60     ip-172-31-8-60.eu-west-1.compute.internal       n2

-- INSERT --
```

- After that you should be able to do ssh n1 to connect to datanode(n1 and n2) from master node(m).



```
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]# exit
logout
Connection to ec2-54-194-24-250.eu-west-1.compute.amazonaws.com closed.
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# vi /etc/hosts
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ssh n1
Warning: Permanently added 'n1' (RSA) to the list of known hosts.
Last login: Sat Nov 16 10:02:40 2013 from 172.31.8.58
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]#
[root@ip-172-31-8-59 ~]# exit
logout
Connection to n1 closed.
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]#
[root@ip-172-31-8-58 ~]# ssh n2
Warning: Permanently added 'n2' (RSA) to the list of known hosts.
Last login: Sat Nov 16 10:00:25 2013 from 172.31.8.58
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
```

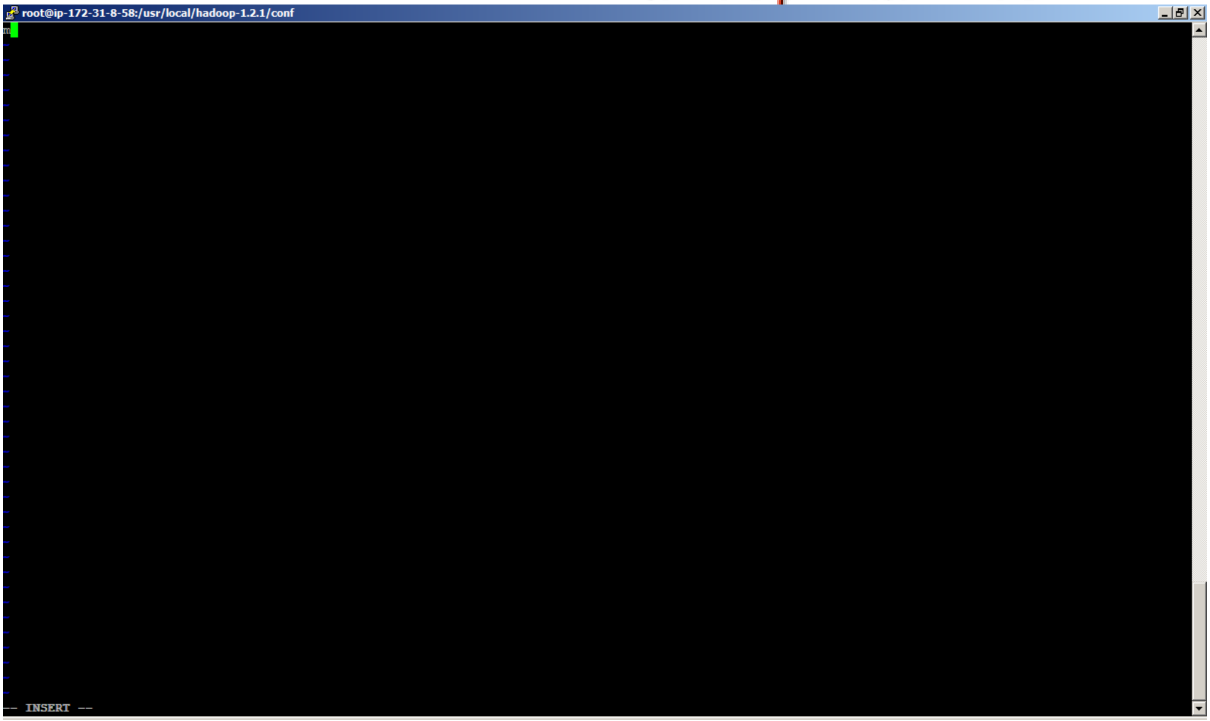
- Now copy the same content in hosts file of remaining nodes n1 and n2

32. Now configure the **Hadoop**

    On Master Node we need to configure the following 5 files to make the three node cluster at path **/usr/local/hadoop-1.2.1/conf**

masters slaves mapred-site.xml hdfs-site.xml core-site.xml

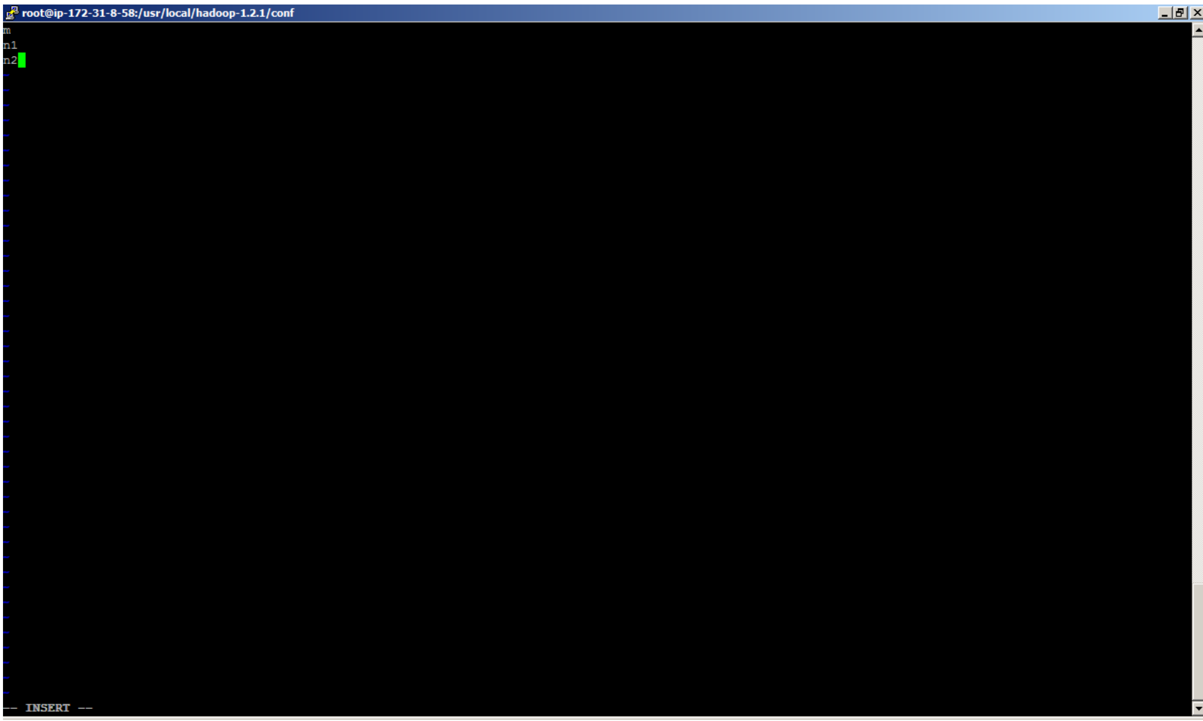
- Open the "masters" file and add the master node(**m**) as below. This file contain the nformation about master nodes.



root@ip-172-31-8-58:/usr/local/hadoop-1.2.1/conf

```
m

-- INSERT --
```

- Now open the slaves file and add both the datanode as well as master node entry as below.

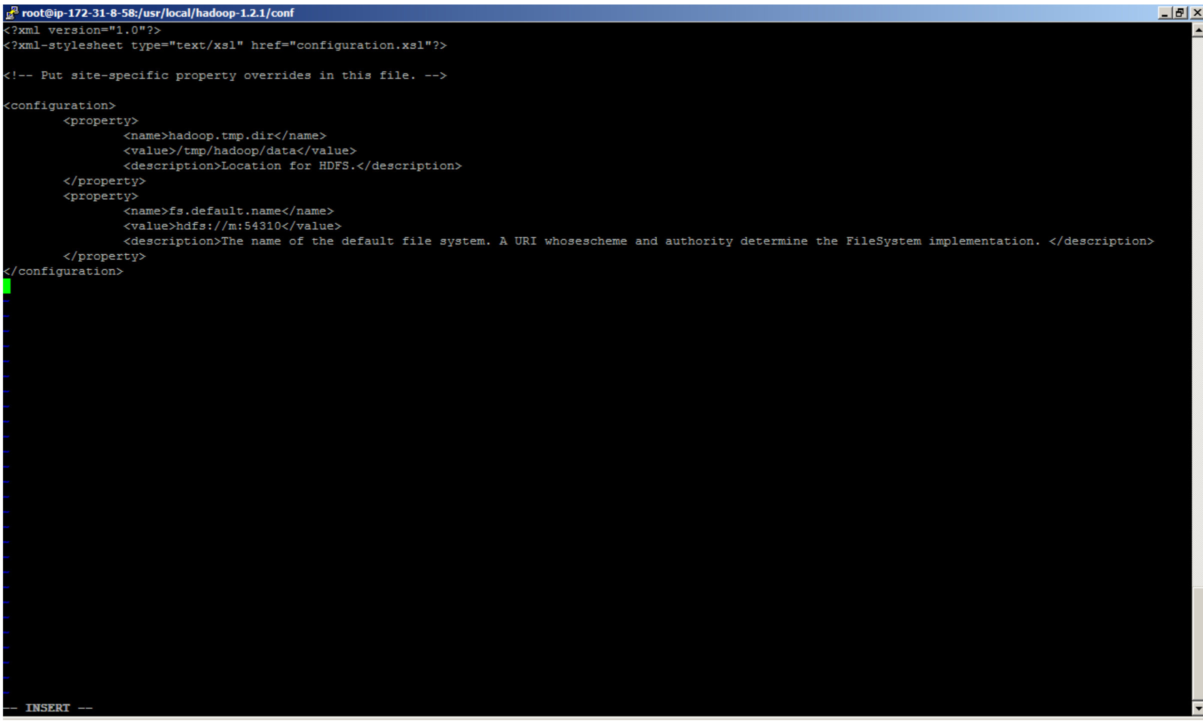


root@ip-172-31-8-58:/usr/local/hadoop-1.2.1/conf

```
m
n1
n2

-- INSERT --
```

- Now configure the “core-site.xml” with following content

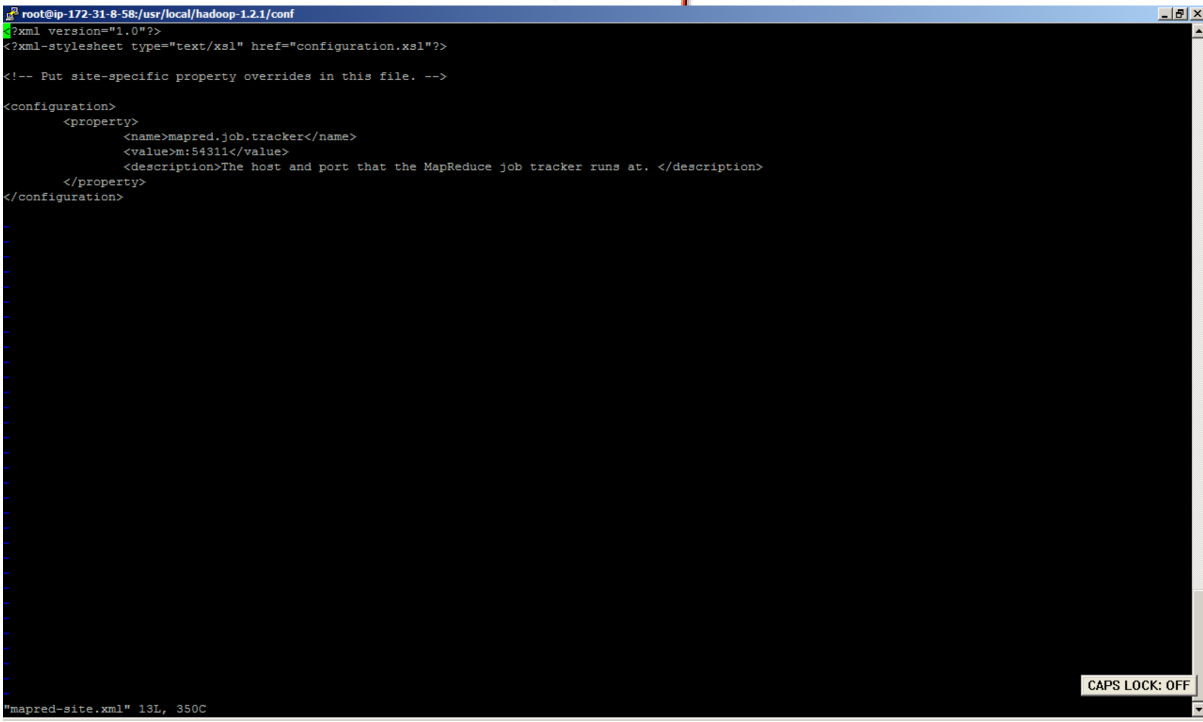
```
<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/tmp/hadoop/data</value>
                <description>Location for HDFS.</description>
        </property>
        <property>
                <name>fs.default.name</name>
                <value>hdfs://m:54310</value>
                <description>The name of the default file system. A URI
whosescheme and authority determine the FileSystem implementation.
</description>
        </property>
</configuration>
```



```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

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

<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/tmp/hadoop/data</value>
                <description>Location for HDFS.</description>
        </property>
        <property>
                <name>fs.default.name</name>
                <value>hdfs://m:54310</value>
                <description>The name of the default file system. A URI whosescheme and authority determine the FileSystem implementation. </description>
        </property>
</configuration>

-- INSERT --
```

- Now configure the mapred-site.xml with following content.

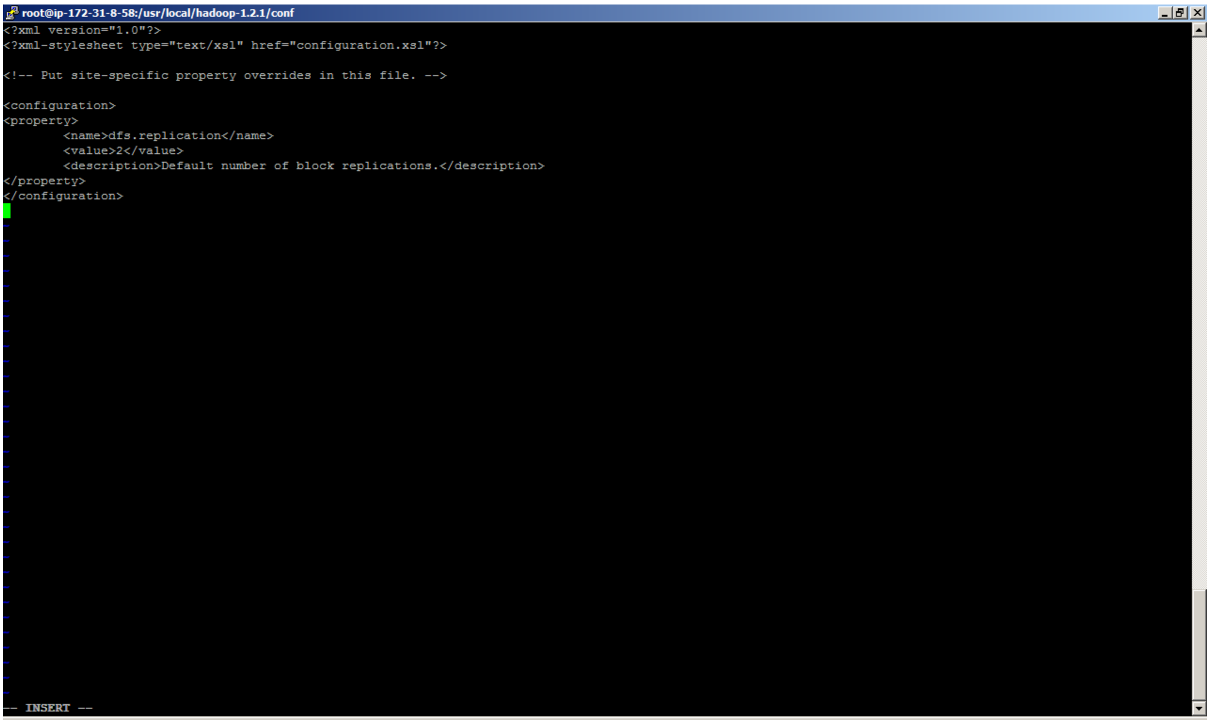
```
<configuration>
        <property>
                <name>mapred.job.tracker</name>
                <value>m:54311</value>
                <description>The host and port that the MapReduce job tracker
runs at. </description>
        </property>
</configuration>
```



```
<?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>m:54311</value>
                <description>The host and port that the MapReduce job tracker runs at. </description>
        </property>
</configuration>

"mapred-site.xml" 13L, 350C
```

- Configure the hdfs-site.xml with following content.

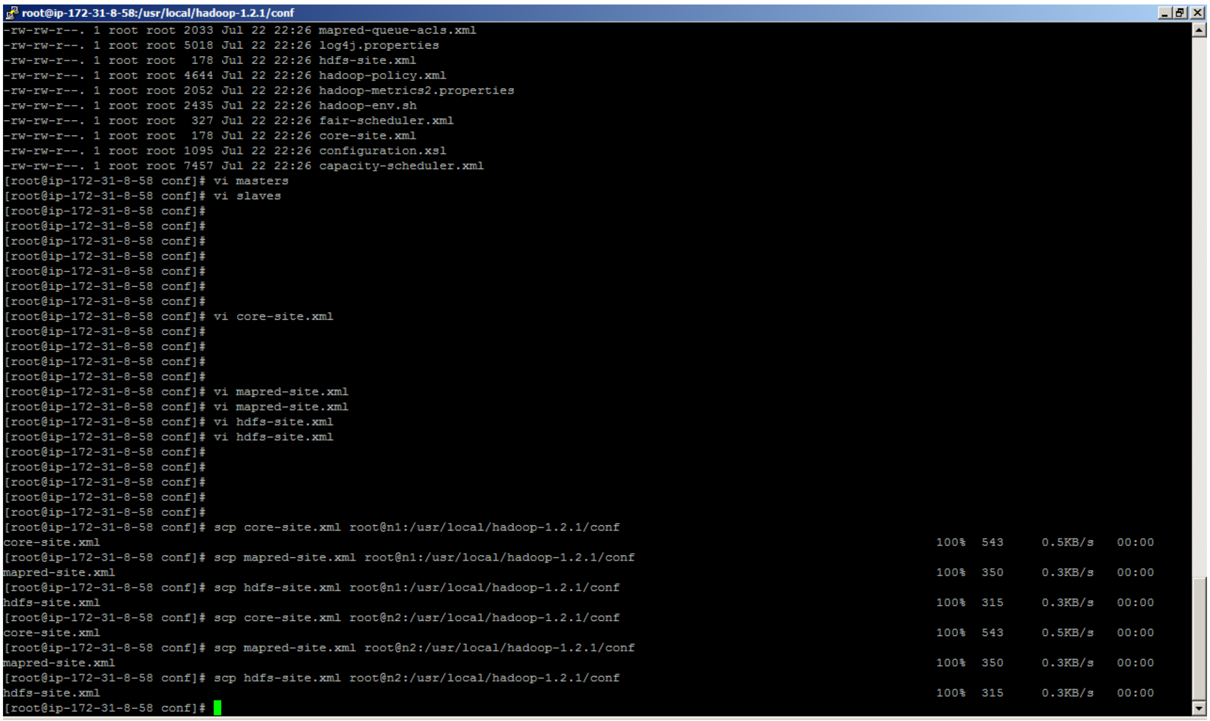
```
<configuration>
<property>
        <name>dfs.replication</name>
        <value>2</value>
        <description>Default number of block replications.</description>
</property>
</configuration>
```



```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

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

<configuration>
<property>
        <name>dfs.replication</name>
        <value>2</value>
        <description>Default number of block replications.</description>
</property>
</configuration>

-- INSERT --
```

33. Now copy the same configuration files on all the datanode as will with following command

```
scp core-site.xml root@n1:/usr/local/hadoop-1.2.1/conf
scp mapred-site.xml root@n1:/usr/local/hadoop-1.2.1/conf
scp hdfs-site.xml root@n1:/usr/local/hadoop-1.2.1/conf

scp core-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
scp mapred-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
scp hdfs-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
```



root@ip-172-31-8-58:/usr/local/hadoop-1.2.1/conf

```
-rw-rw-r--. 1 root root 2033 Jul 22 22:26 mapred-queue-acls.xml
-rw-rw-r--. 1 root root 5018 Jul 22 22:26 log4j.properties
-rw-rw-r--. 1 root root  178 Jul 22 22:26 hdfs-site.xml
-rw-rw-r--. 1 root root 4644 Jul 22 22:26 hadoop-policy.xml
-rw-rw-r--. 1 root root 2052 Jul 22 22:26 hadoop-metrics2.properties
-rw-rw-r--. 1 root root 2435 Jul 22 22:26 hadoop-env.sh
-rw-rw-r--. 1 root root  327 Jul 22 22:26 fair-scheduler.xml
-rw-rw-r--. 1 root root  178 Jul 22 22:26 core-site.xml
-rw-rw-r--. 1 root root 1095 Jul 22 22:26 configuration.xsl
-rw-rw-r--. 1 root root 7457 Jul 22 22:26 capacity-scheduler.xml
[root@ip-172-31-8-58 conf]# vi masters
[root@ip-172-31-8-58 conf]# vi slaves
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]# vi core-site.xml
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]# vi mapred-site.xml
[root@ip-172-31-8-58 conf]# vi mapred-site.xml
[root@ip-172-31-8-58 conf]# vi hdfs-site.xml
[root@ip-172-31-8-58 conf]# vi hdfs-site.xml
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]# scp core-site.xml root@n1:/usr/local/hadoop-1.2.1/conf
core-site.xml                                                                 100%  543     0.5KB/s   00:00
[root@ip-172-31-8-58 conf]# scp mapred-site.xml root@n1:/usr/local/hadoop-1.2.1/conf
mapred-site.xml                                                               100%  350     0.3KB/s   00:00
[root@ip-172-31-8-58 conf]# scp hdfs-site.xml root@n1:/usr/local/hadoop-1.2.1/conf
hdfs-site.xml                                                                 100%  315     0.3KB/s   00:00
[root@ip-172-31-8-58 conf]# scp core-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
core-site.xml                                                                 100%  543     0.5KB/s   00:00
[root@ip-172-31-8-58 conf]# scp mapred-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
mapred-site.xml                                                               100%  350     0.3KB/s   00:00
[root@ip-172-31-8-58 conf]# scp hdfs-site.xml root@n2:/usr/local/hadoop-1.2.1/conf
hdfs-site.xml                                                                 100%  315     0.3KB/s   00:00
[root@ip-172-31-8-58 conf]#
```

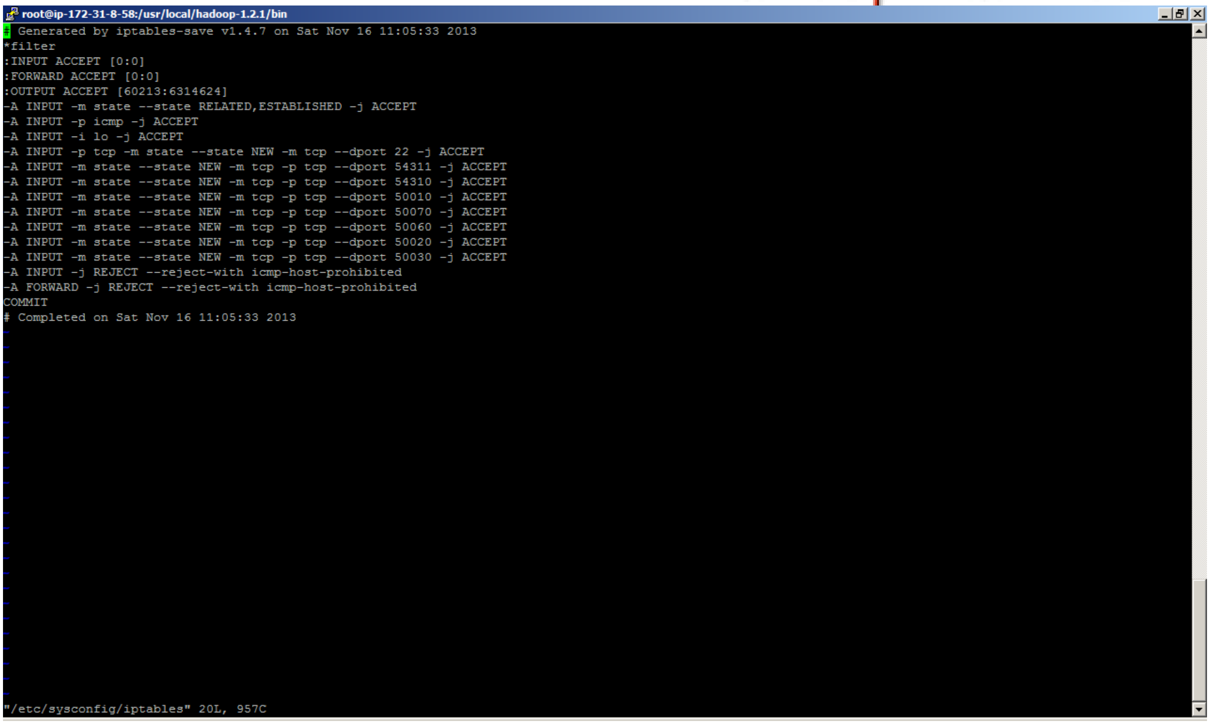
34. Now we need to do firewall settings so that each node can also access each other using the various required port. We need to do following configuration on each node.

Open the file iptables file from following command.

```
vi /etc/sysconfig/iptables
```

And add the following entry just below the Port 22 configuration in the files.

```
-A INPUT -m state --state NEW -m tcp -p tcp --dport 54311 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 54310 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50010 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50070 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50060 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50020 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50030 -j ACCEPT
```



root@ip-172-31-8-58:/usr/local/hadoop-1.2.1/bin

```
# Generated by iptables-save v1.4.7 on Sat Nov 16 11:05:33 2013
*filter
:INPUT ACCEPT [0:0]
:FORWARD ACCEPT [0:0]
:OUTPUT ACCEPT [60213:6314624]
-A INPUT -m state --state RELATED,ESTABLISHED -j ACCEPT
-A INPUT -p icmp -j ACCEPT
-A INPUT -i lo -j ACCEPT
-A INPUT -p tcp -m state --state NEW -m tcp --dport 22 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 54311 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 54310 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50010 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50070 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50060 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50020 -j ACCEPT
-A INPUT -m state --state NEW -m tcp -p tcp --dport 50030 -j ACCEPT
-A INPUT -j REJECT --reject-with icmp-host-prohibited
-A FORWARD -j REJECT --reject-with icmp-host-prohibited
COMMIT
# Completed on Sat Nov 16 11:05:33 2013

"/etc/sysconfig/iptables" 20L, 957C
```

Login as a root to all nodes execute the following commands.

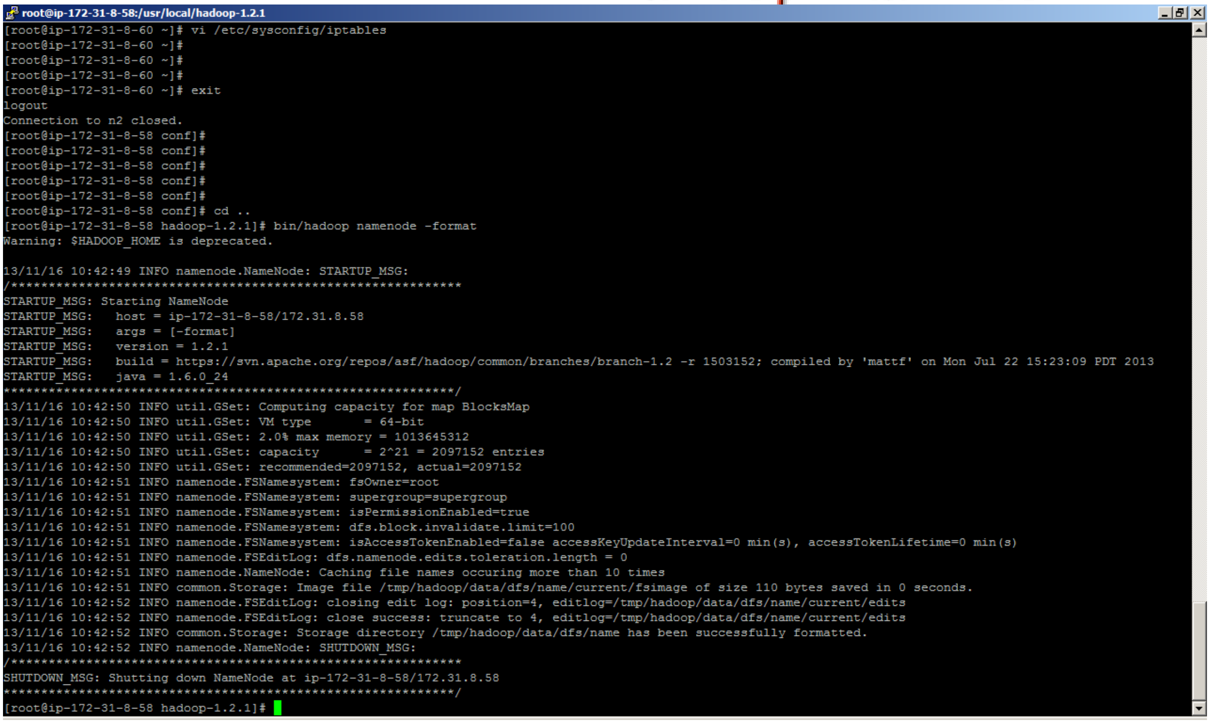
```
# service iptables save
# service iptables stop
# chkconfig iptables off
```

Please do the same entries on all the nodes iptables.

Start Hadoop Cluster

35. Format the namenode to create HDFS filesystem. Formatting the Hadoop filesystem, which is implemented on top of the local filesystems of your cluster, you need to do this the first time you set up a Hadoop installation. Do not format a running Hadoop filesystem, this will cause all your data to be erased.

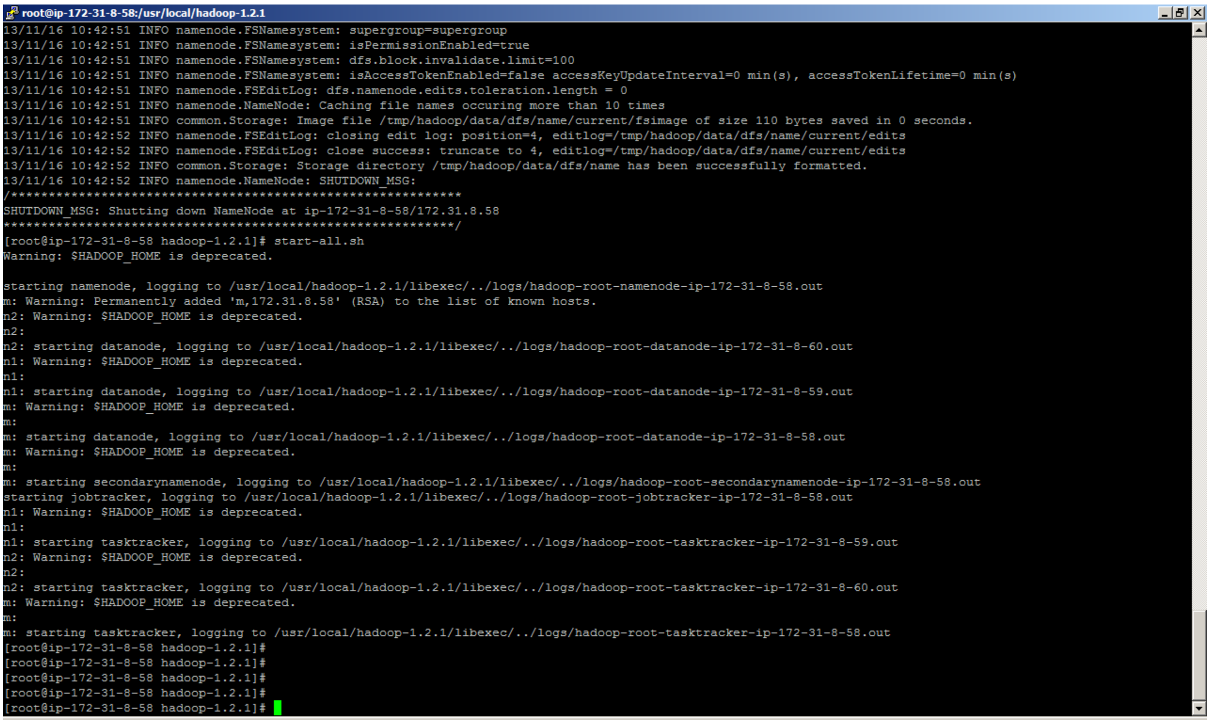
```
bin/hadoop namenode -format
```



```
[root@ip-172-31-8-60 ~]# vi /etc/sysconfig/iptables
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]#
[root@ip-172-31-8-60 ~]# exit
logout
Connection to n2 closed.
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]#
[root@ip-172-31-8-58 conf]# cd ..
[root@ip-172-31-8-58 hadoop-1.2.1]# bin/hadoop namenode -format
Warning: $HADOOP_HOME is deprecated.

13/11/16 10:42:49 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = ip-172-31-8-58/172.31.8.58
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/branch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.6.0_24
************************************************************/
13/11/16 10:42:50 INFO util.GSet: Computing capacity for map BlocksMap
13/11/16 10:42:50 INFO util.GSet: VM type       = 64-bit
13/11/16 10:42:50 INFO util.GSet: 2.0% max memory = 1013645312
13/11/16 10:42:50 INFO util.GSet: capacity      = 2^21 = 2097152 entries
13/11/16 10:42:50 INFO util.GSet: recommended=2097152, actual=2097152
13/11/16 10:42:51 INFO namenode.FSNamesystem: fsOwner=root
13/11/16 10:42:51 INFO namenode.FSNamesystem: supergroup=supergroup
13/11/16 10:42:51 INFO namenode.FSNamesystem: isPermissionEnabled=true
13/11/16 10:42:51 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
13/11/16 10:42:51 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterval=0 min(s), accessTokenLifetime=0 min(s)
13/11/16 10:42:51 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
13/11/16 10:42:51 INFO namenode.NameNode: Caching file names occuring more than 10 times
13/11/16 10:42:51 INFO common.Storage: Image file /tmp/hadoop/data/dfs/name/current/fsimage of size 110 bytes saved in 0 seconds.
13/11/16 10:42:52 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/tmp/hadoop/data/dfs/name/current/edits
13/11/16 10:42:52 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/tmp/hadoop/data/dfs/name/current/edits
13/11/16 10:42:52 INFO common.Storage: Storage directory /tmp/hadoop/data/dfs/name has been successfully formatted.
13/11/16 10:42:52 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-8-58/172.31.8.58
************************************************************/
[root@ip-172-31-8-58 hadoop-1.2.1]#
```

36. Now it’s time to start your Hadoop Multi Node Cluster with following command on master node.

```
start-all.sh
```



```
13/11/16 10:42:51 INFO namenode.FSNamesystem: supergroup=supergroup
13/11/16 10:42:51 INFO namenode.FSNamesystem: isPermissionEnabled=true
13/11/16 10:42:51 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
13/11/16 10:42:51 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterval=0 min(s), accessTokenLifetime=0 min(s)
13/11/16 10:42:51 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
13/11/16 10:42:51 INFO namenode.NameNode: Caching file names occuring more than 10 times
13/11/16 10:42:51 INFO common.Storage: Image file /tmp/hadoop/data/dfs/name/current/fsimage of size 110 bytes saved in 0 seconds.
13/11/16 10:42:52 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/tmp/hadoop/data/dfs/name/current/edits
13/11/16 10:42:52 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/tmp/hadoop/data/dfs/name/current/edits
13/11/16 10:42:52 INFO common.Storage: Storage directory /tmp/hadoop/data/dfs/name has been successfully formatted.
13/11/16 10:42:52 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-8-58/172.31.8.58
************************************************************/
[root@ip-172-31-8-58 hadoop-1.2.1]# start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-namenode-ip-172-31-8-58.out
m: Warning: Permanently added 'm,172.31.8.58' (RSA) to the list of known hosts.
n2: Warning: $HADOOP_HOME is deprecated.
n2:
n2: starting datanode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-datanode-ip-172-31-8-60.out
n1: Warning: $HADOOP_HOME is deprecated.
n1:
n1: starting datanode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-datanode-ip-172-31-8-59.out
m: Warning: $HADOOP_HOME is deprecated.
m:
m: starting datanode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-datanode-ip-172-31-8-58.out
m: Warning: $HADOOP_HOME is deprecated.
m:
m: starting secondarynamenode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-secondarynamenode-ip-172-31-8-58.out
starting jobtracker, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-jobtracker-ip-172-31-8-58.out
n1: Warning: $HADOOP_HOME is deprecated.
n1:
n1: starting tasktracker, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-tasktracker-ip-172-31-8-59.out
n2: Warning: $HADOOP_HOME is deprecated.
n2:
n2: starting tasktracker, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-tasktracker-ip-172-31-8-60.out
m: Warning: $HADOOP_HOME is deprecated.
m:
m: starting tasktracker, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-root-tasktracker-ip-172-31-8-58.out
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
```

37. Now to verify we should whether our cluster started properly or not. Using the following command it should list all the java processes. Repeat verification on all the nodes.

```
Ps –aef | grep java
```

It will list all the process of daemon.

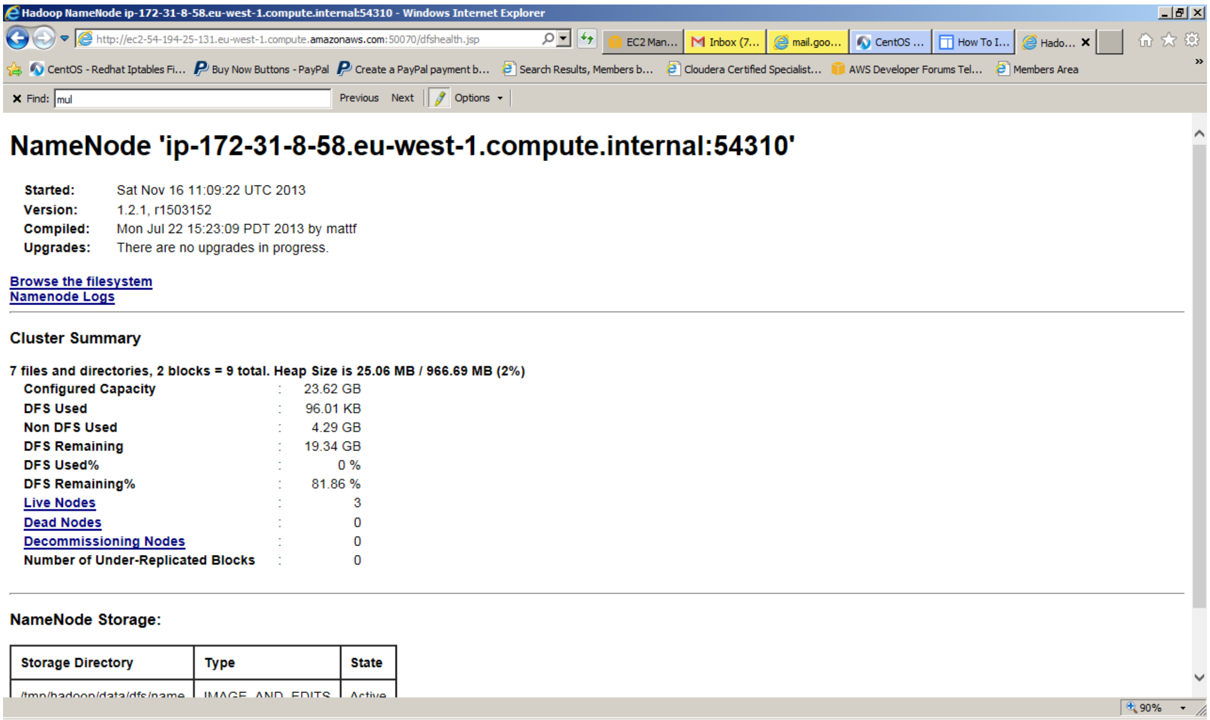
On master node there should be all 5 daemons (NameNode, SecondaryNameNode, DataNode, JobTracker and TaskTracker should be running) and on data node you see (JobTracker and Tasktracker)

Output shown in video trainings, please check.

38. Check the NameNode UI as below.

    http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50070

    It is showing two live nodes.



Hadoop NameNode ip-172-31-8-58.eu-west-1.compute.internal:54310 - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50070/dfshealth.jsp
EC2 Man...
Inbox (7...
mail.goo...
CentOS ...
How To I...
Hado...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Find: mul
Previous
Next
Options
90%

NameNode 'ip-172-31-8-58.eu-west-1.compute.internal:54310'

Started:	Sat Nov 16 11:09:22 UTC 2013
Version:	1.2.1, r1503152
Compiled:	Mon Jul 22 15:23:09 PDT 2013 by mattf
Upgrades:	There are no upgrades in progress.

Browse the filesystem
Namenode Logs

Cluster Summary

7 files and directories, 2 blocks = 9 total. Heap Size is 25.06 MB / 966.69 MB (2%)

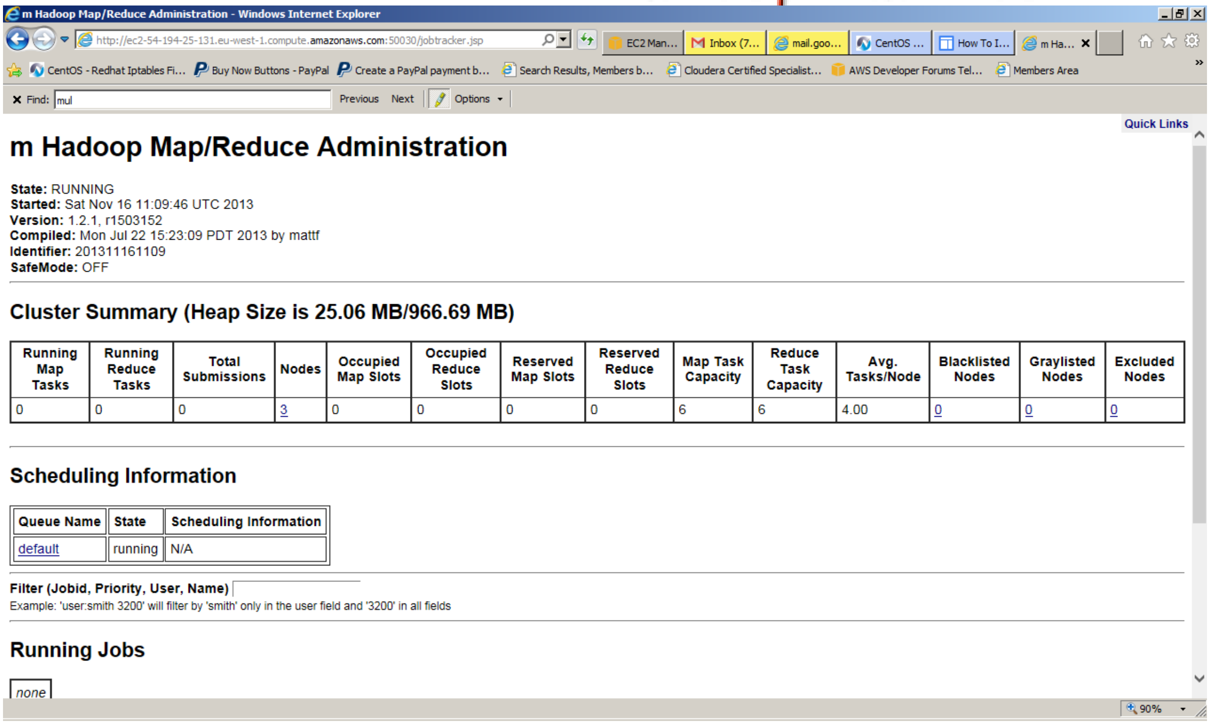
Configured Capacity	:	23.62 GB
DFS Used	:	96.01 KB
Non DFS Used	:	4.29 GB
DFS Remaining	:	19.34 GB
DFS Used%	:	0 %
DFS Remaining%	:	81.86 %
Live Nodes	:	3
Dead Nodes	:	0
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

NameNode Storage:

Storage Directory	Type	State
/tmp/hadoop/data/dfs/name	IMAGE_AND_EDITS	Active

39. JobTracker UI (We have not yet executed any job so it shows 0 jobs)

    http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030



m Hadoop Map/Reduce Administration - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/jobtracker.jsp
Find: mul
Previous
Next
Options
Quick Links
90%

m Hadoop Map/Reduce Administration

State: RUNNING
Started: Sat Nov 16 11:09:46 UTC 2013
Version: 1.2.1, r1503152
Compiled: Mon Jul 22 15:23:09 PDT 2013 by mattf
Identifier: 201311161109
SafeMode: OFF

Cluster Summary (Heap Size is 25.06 MB/966.69 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Graylisted Nodes	Excluded Nodes
0	0	0	3	0	0	0	0	6	6	4.00	0	0	0

Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields

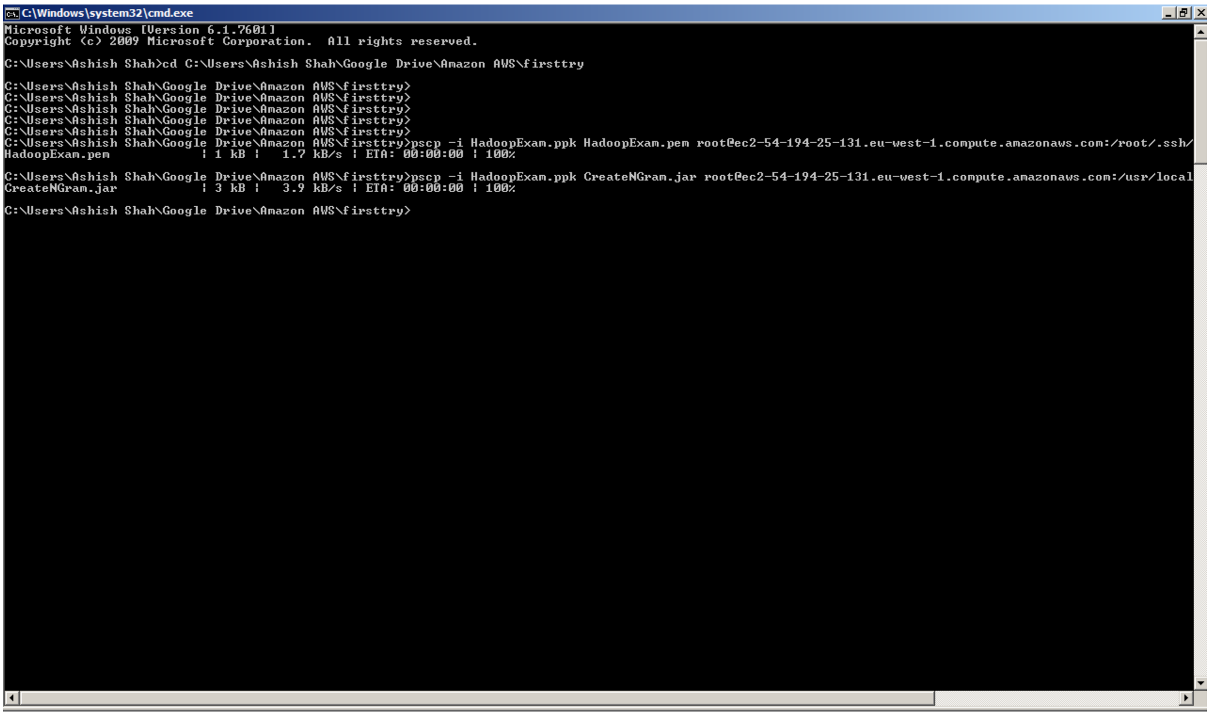
Running Jobs

none

40. Running the example , In module 15 we have created **NGram** example same we will run on this cluster.

    - Copy the code/jar files on Hadoop Master Node using following code.

```
pscp -i HadoopExam.ppk CreateNGram.jar root@ec2-54-194-25-131.eu-west-
1.compute.amazonaws.com:/usr/local/hadoop-1.2.1
```



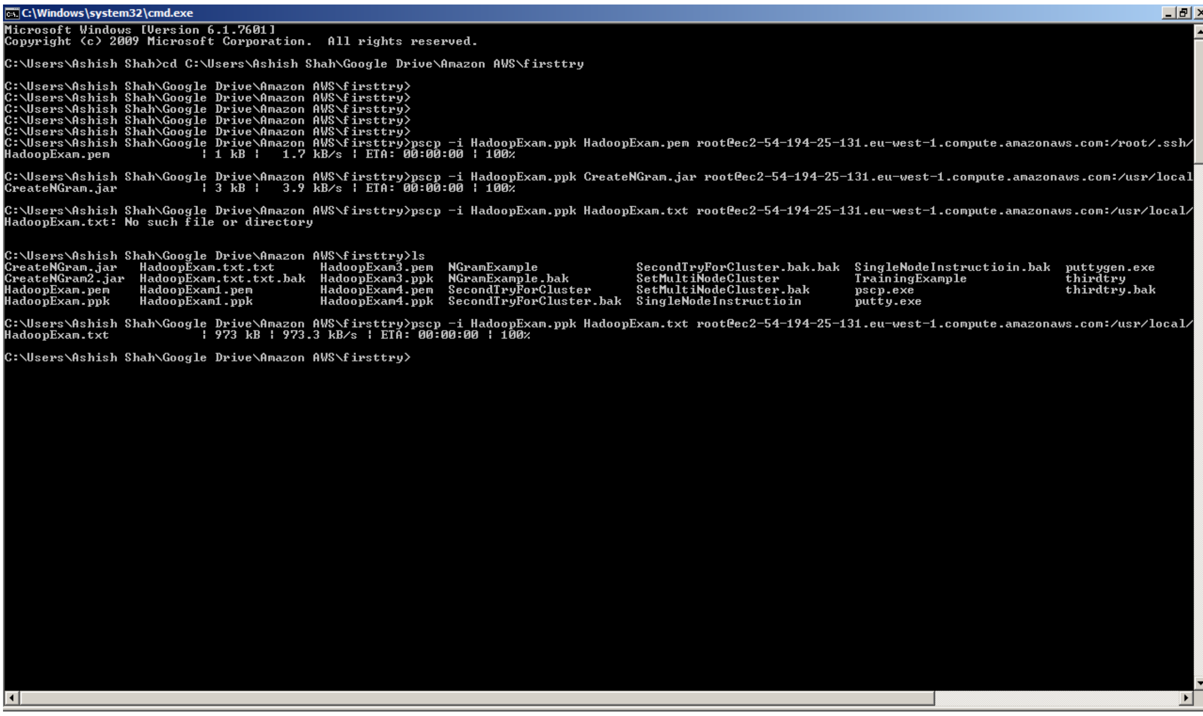
```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Ashish Shah>cd C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.pem root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/root/.ssh/
HadoopExam.pem            | 1 kB |   1.7 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk CreateNGram.jar root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/usr/local
CreateNGram.jar           | 3 kB |   3.9 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
```

- Now create the Temporary text file(Little bigger) on local machine and then copy it to HDFS.

```
mkdir /usr/local/tempData
```

- And copy a big text file from your local machine to Master Node local directory.

```
pscp -i HadoopExam.ppk HadoopExam.txt root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/usr/local/tempData/
```



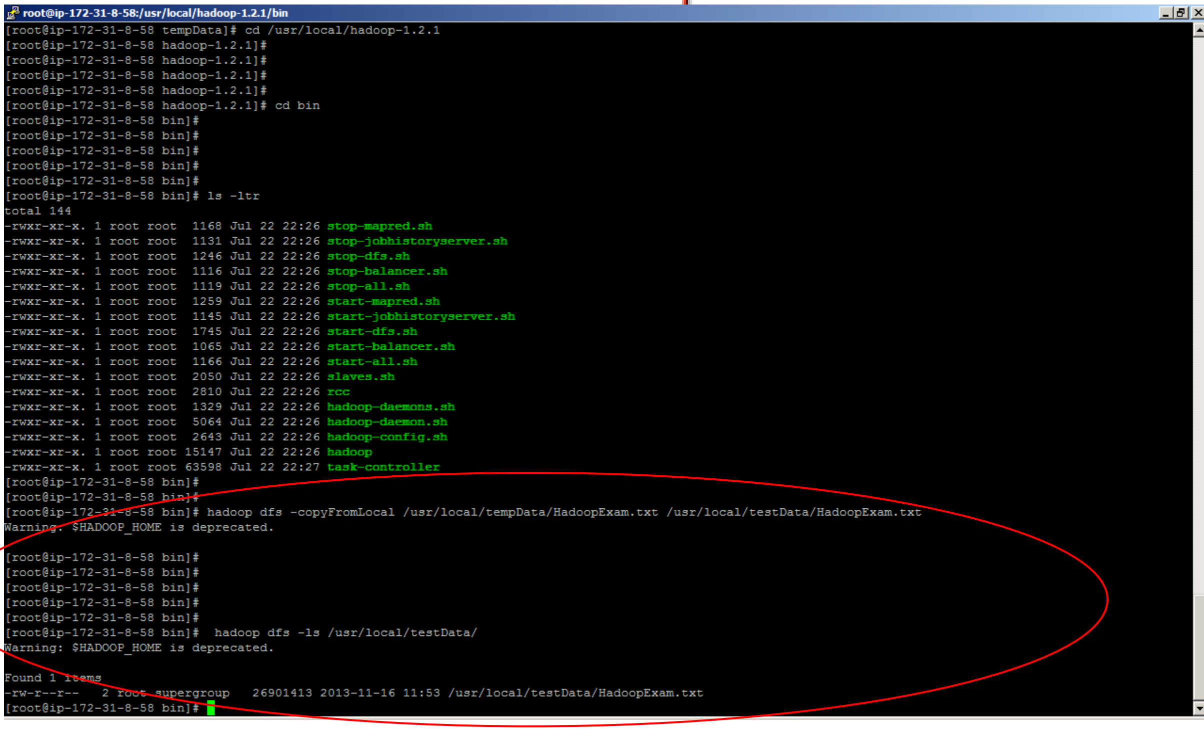
```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Ashish Shah>cd C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.pem root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/root/.ssh/
HadoopExam.pem            | 1 kB |   1.7 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk CreateNGram.jar root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/usr/local
CreateNGram.jar           | 3 kB |   3.9 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.txt root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/usr/local/
HadoopExam.txt: No such file or directory


C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>ls
CreateNGram.jar   HadoopExam.txt.txt      HadoopExam3.pem  NGramExample             SecondTryForCluster.bak.bak  SingleNodeInstructioin.bak  puttygen.exe
CreateNGram2.jar  HadoopExam.txt.txt.bak  HadoopExam3.ppk  NGramExample.bak         SetMultiNodeCluster          TrainingExample             thirdtry
HadoopExam.pem    HadoopExam1.pem         HadoopExam4.pem  SecondTryForCluster      SetMultiNodeCluster.bak      pscp.exe                    thirdtry.bak
HadoopExam.ppk    HadoopExam1.ppk         HadoopExam4.ppk  SecondTryForCluster.bak  SingleNodeInstructioin       putty.exe

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>pscp -i HadoopExam.ppk HadoopExam.txt root@ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:/usr/local/
HadoopExam.txt            | 973 kB | 973.3 kB/s | ETA: 00:00:00 | 100%

C:\Users\Ashish Shah\Google Drive\Amazon AWS\firsttry>
```

- Now copy this bigfile from local disk of master node(Multiple times) to HDFS file system.

```
hadoop dfs -copyFromLocal /usr/local/tempData/HadoopExam.txt /usr/local/testData/HadoopExam.txt

hadoop dfs -copyFromLocal /usr/local/tempData/HadoopExam.txt /usr/local/testData/HadoopExam1.txt

hadoop dfs -copyFromLocal /usr/local/tempData/HadoopExam.txt /usr/local/testData/HadoopExam2.txt

hadoop dfs -copyFromLocal /usr/local/tempData/HadoopExam.txt /usr/local/testData/HadoopExam3.txt
```

- Check whether file has been created or not.

```
bin/hadoop dfs -ls /usr/local/testData
```



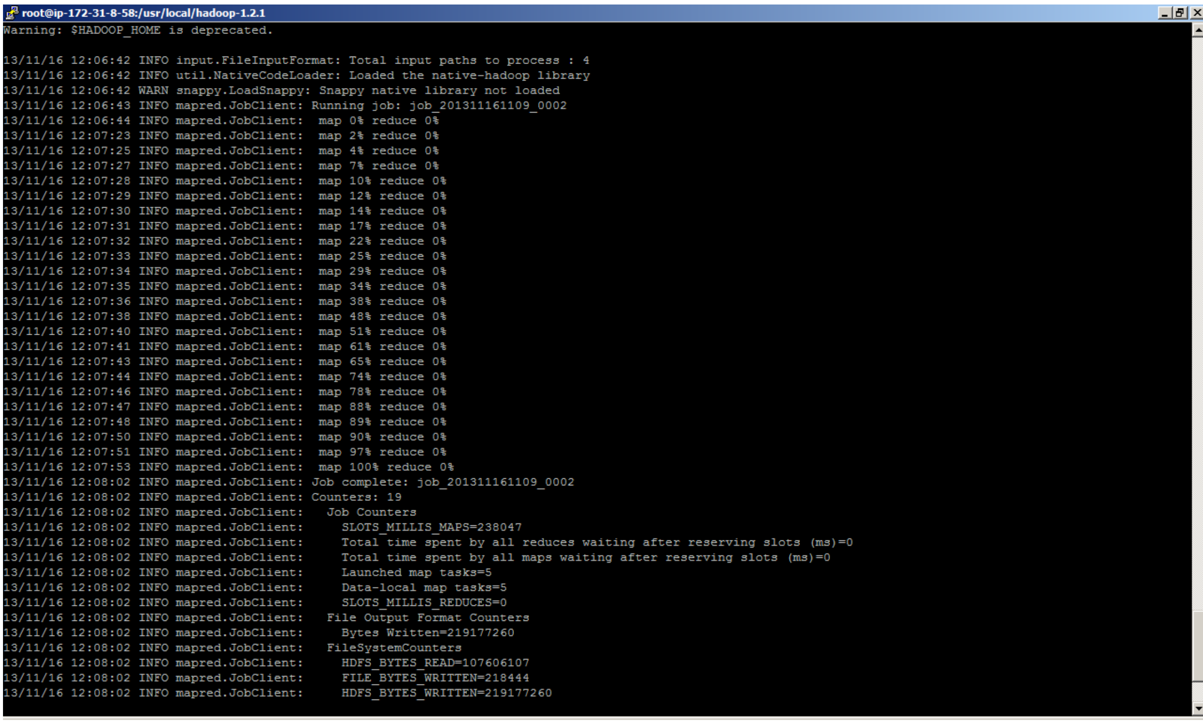
root@ip-172-31-8-58:/usr/local/hadoop-1.2.1/bin

```
[root@ip-172-31-8-58 tempData]# cd /usr/local/hadoop-1.2.1
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]#
[root@ip-172-31-8-58 hadoop-1.2.1]# cd bin
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]# ls -ltr
total 144
-rwxr-xr-x. 1 root root  1168 Jul 22 22:26 stop-mapred.sh
-rwxr-xr-x. 1 root root  1131 Jul 22 22:26 stop-jobhistoryserver.sh
-rwxr-xr-x. 1 root root  1246 Jul 22 22:26 stop-dfs.sh
-rwxr-xr-x. 1 root root  1116 Jul 22 22:26 stop-balancer.sh
-rwxr-xr-x. 1 root root  1119 Jul 22 22:26 stop-all.sh
-rwxr-xr-x. 1 root root  1259 Jul 22 22:26 start-mapred.sh
-rwxr-xr-x. 1 root root  1145 Jul 22 22:26 start-jobhistoryserver.sh
-rwxr-xr-x. 1 root root  1745 Jul 22 22:26 start-dfs.sh
-rwxr-xr-x. 1 root root  1065 Jul 22 22:26 start-balancer.sh
-rwxr-xr-x. 1 root root  1166 Jul 22 22:26 start-all.sh
-rwxr-xr-x. 1 root root  2050 Jul 22 22:26 slaves.sh
-rwxr-xr-x. 1 root root  2810 Jul 22 22:26 rcc
-rwxr-xr-x. 1 root root  1329 Jul 22 22:26 hadoop-daemons.sh
-rwxr-xr-x. 1 root root  5064 Jul 22 22:26 hadoop-daemon.sh
-rwxr-xr-x. 1 root root  2643 Jul 22 22:26 hadoop-config.sh
-rwxr-xr-x. 1 root root 15147 Jul 22 22:26 hadoop
-rwxr-xr-x. 1 root root 63598 Jul 22 22:27 task-controller
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]# hadoop dfs -copyFromLocal /usr/local/tempData/HadoopExam.txt /usr/local/testData/HadoopExam.txt
Warning: $HADOOP_HOME is deprecated.

[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#
[root@ip-172-31-8-58 bin]#  hadoop dfs -ls /usr/local/testData/
Warning: $HADOOP_HOME is deprecated.

Found 1 items
-rw-r--r--   2 root supergroup   26901413 2013-11-16 11:53 /usr/local/testData/HadoopExam.txt
[root@ip-172-31-8-58 bin]#
```

- Run the program

```
bin/hadoop jar CreateNGram.jar com.hadoop.exam.CreateNGramDriver
/usr/local/testData/ /usr/local/testData-output1 3
```

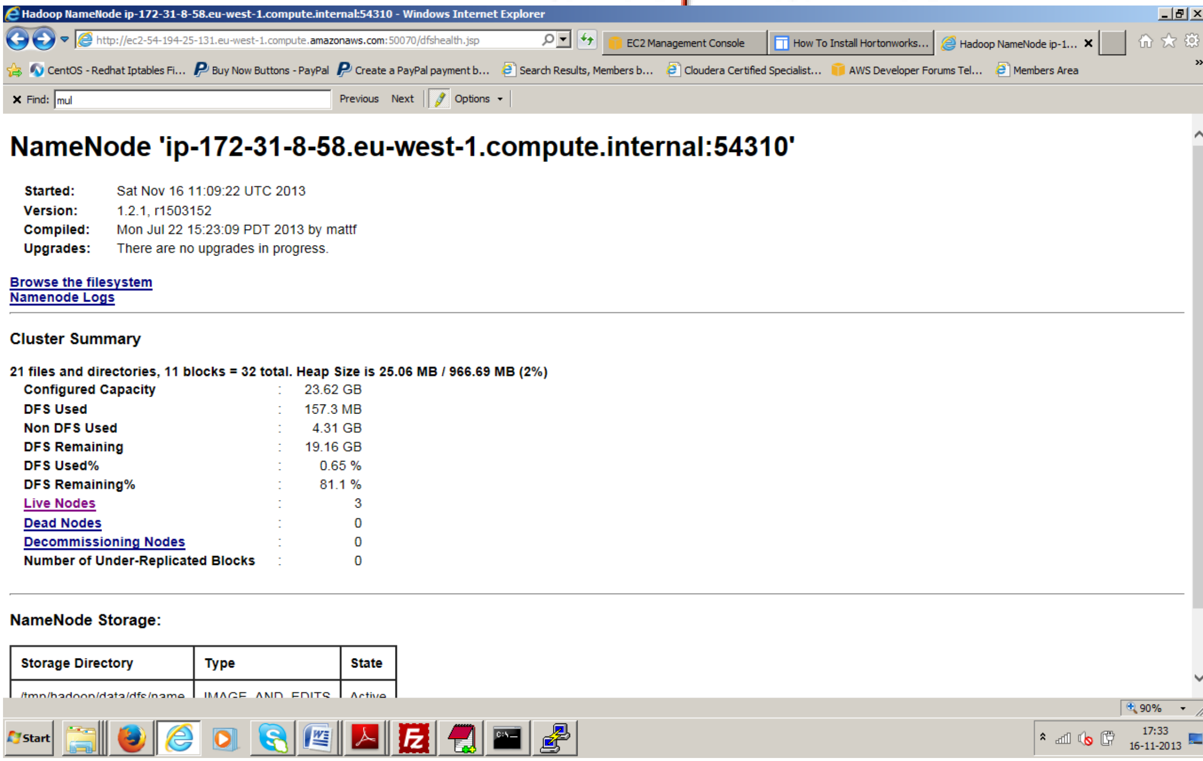
- Ouput of the Job on Console



root@ip-172-31-8-58:/usr/local/hadoop-1.2.1

```
Warning: $HADOOP_HOME is deprecated.

13/11/16 12:06:42 INFO input.FileInputFormat: Total input paths to process : 4
13/11/16 12:06:42 INFO util.NativeCodeLoader: Loaded the native-hadoop library
13/11/16 12:06:42 WARN snappy.LoadSnappy: Snappy native library not loaded
13/11/16 12:06:43 INFO mapred.JobClient: Running job: job_201311161109_0002
13/11/16 12:06:44 INFO mapred.JobClient:  map 0% reduce 0%
13/11/16 12:07:23 INFO mapred.JobClient:  map 2% reduce 0%
13/11/16 12:07:25 INFO mapred.JobClient:  map 4% reduce 0%
13/11/16 12:07:27 INFO mapred.JobClient:  map 7% reduce 0%
13/11/16 12:07:28 INFO mapred.JobClient:  map 10% reduce 0%
13/11/16 12:07:29 INFO mapred.JobClient:  map 12% reduce 0%
13/11/16 12:07:30 INFO mapred.JobClient:  map 14% reduce 0%
13/11/16 12:07:31 INFO mapred.JobClient:  map 17% reduce 0%
13/11/16 12:07:32 INFO mapred.JobClient:  map 22% reduce 0%
13/11/16 12:07:33 INFO mapred.JobClient:  map 25% reduce 0%
13/11/16 12:07:34 INFO mapred.JobClient:  map 29% reduce 0%
13/11/16 12:07:35 INFO mapred.JobClient:  map 34% reduce 0%
13/11/16 12:07:36 INFO mapred.JobClient:  map 38% reduce 0%
13/11/16 12:07:38 INFO mapred.JobClient:  map 48% reduce 0%
13/11/16 12:07:40 INFO mapred.JobClient:  map 51% reduce 0%
13/11/16 12:07:41 INFO mapred.JobClient:  map 61% reduce 0%
13/11/16 12:07:43 INFO mapred.JobClient:  map 65% reduce 0%
13/11/16 12:07:44 INFO mapred.JobClient:  map 74% reduce 0%
13/11/16 12:07:46 INFO mapred.JobClient:  map 78% reduce 0%
13/11/16 12:07:47 INFO mapred.JobClient:  map 88% reduce 0%
13/11/16 12:07:48 INFO mapred.JobClient:  map 89% reduce 0%
13/11/16 12:07:50 INFO mapred.JobClient:  map 90% reduce 0%
13/11/16 12:07:51 INFO mapred.JobClient:  map 97% reduce 0%
13/11/16 12:07:53 INFO mapred.JobClient:  map 100% reduce 0%
13/11/16 12:08:02 INFO mapred.JobClient: Job complete: job_201311161109_0002
13/11/16 12:08:02 INFO mapred.JobClient: Counters: 19
13/11/16 12:08:02 INFO mapred.JobClient:   Job Counters
13/11/16 12:08:02 INFO mapred.JobClient:     SLOTS_MILLIS_MAPS=238047
13/11/16 12:08:02 INFO mapred.JobClient:     Total time spent by all reduces waiting after reserving slots (ms)=0
13/11/16 12:08:02 INFO mapred.JobClient:     Total time spent by all maps waiting after reserving slots (ms)=0
13/11/16 12:08:02 INFO mapred.JobClient:     Launched map tasks=5
13/11/16 12:08:02 INFO mapred.JobClient:     Data-local map tasks=5
13/11/16 12:08:02 INFO mapred.JobClient:     SLOTS_MILLIS_REDUCES=0
13/11/16 12:08:02 INFO mapred.JobClient:   File Output Format Counters
13/11/16 12:08:02 INFO mapred.JobClient:     Bytes Written=219177260
13/11/16 12:08:02 INFO mapred.JobClient:   FileSystemCounters
13/11/16 12:08:02 INFO mapred.JobClient:     HDFS_BYTES_READ=107606107
13/11/16 12:08:02 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=218444
13/11/16 12:08:02 INFO mapred.JobClient:     HDFS_BYTES_WRITTEN=219177260
```

- After Running the Job Below is the NameNode UI status, number of blocks increased. As new data has been copies.



NameNode 'ip-172-31-8-58.eu-west-1.compute.internal:54310'

Started:	Sat Nov 16 11:09:22 UTC 2013
Version:	1.2.1, r1503152
Compiled:	Mon Jul 22 15:23:09 PDT 2013 by mattf
Upgrades:	There are no upgrades in progress.

Browse the filesystem
Namenode Logs

Cluster Summary

21 files and directories, 11 blocks = 32 total. Heap Size is 25.06 MB / 966.69 MB (2%)

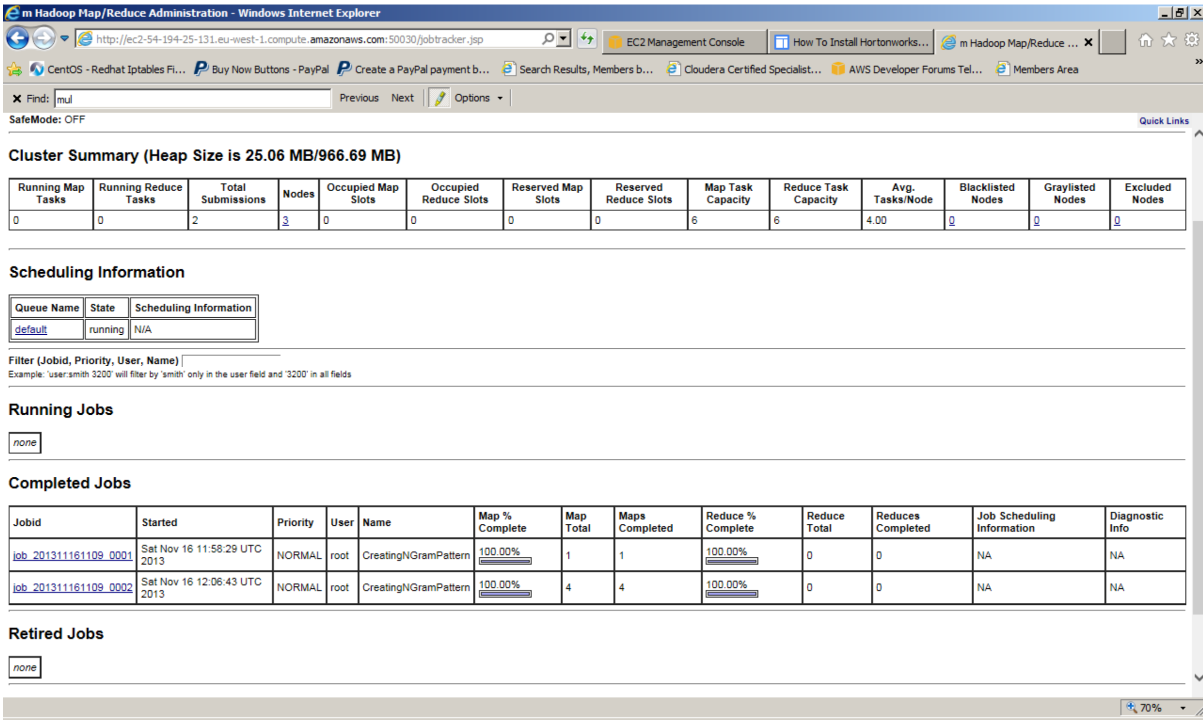
Configured Capacity	:	23.62 GB
DFS Used	:	157.3 MB
Non DFS Used	:	4.31 GB
DFS Remaining	:	19.16 GB
DFS Used%	:	0.65 %
DFS Remaining%	:	81.1 %
Live Nodes	:	3
Dead Nodes	:	0
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

NameNode Storage:

Storage Directory	Type	State
/tmp/hadoop/data/dfs/name	IMAGE_AND_EDITS	Active

Hadoop NameNode ip-172-31-8-58.eu-west-1.compute.internal:54310 - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50070/dfshealth.jsp
EC2 Management Console
How To Install Hortonworks...
Hadoop NameNode ip-1...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Find: mul
Previous
Next
Options
90%
Start
17:33
16-11-2013

- Check JobTracker UI (Two sample jobs we have executed as below)



SafeMode: OFF

Cluster Summary (Heap Size is 25.06 MB/966.69 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Graylisted Nodes	Excluded Nodes
0	0	2	3	0	0	0	0	6	6	4.00	0	0	0

Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields

Running Jobs

none

Completed Jobs

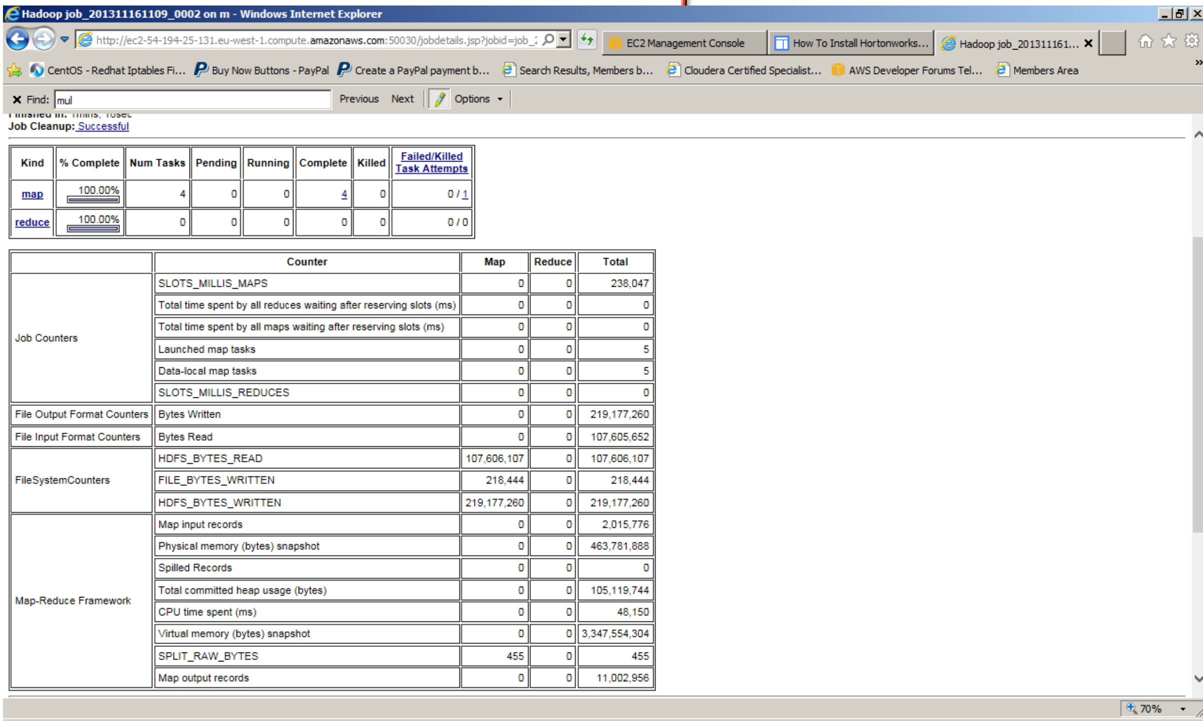
Jobid	Started	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
job_201311161109_0001	Sat Nov 16 11:58:29 UTC 2013	NORMAL	root	CreatingNGramPattern	100.00%	1	1	100.00%	0	0	NA	NA
job_201311161109_0002	Sat Nov 16 12:06:43 UTC 2013	NORMAL	root	CreatingNGramPattern	100.00%	4	4	100.00%	0	0	NA	NA

Retired Jobs

none

m Hadoop Map/Reduce Administration - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/jobtracker.jsp
EC2 Management Console
How To Install Hortonworks...
m Hadoop Map/Reduce ...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Find: mul
Previous
Next
Options
Quick Links
70%

- Click on Second Job which has 4 Map Tasks



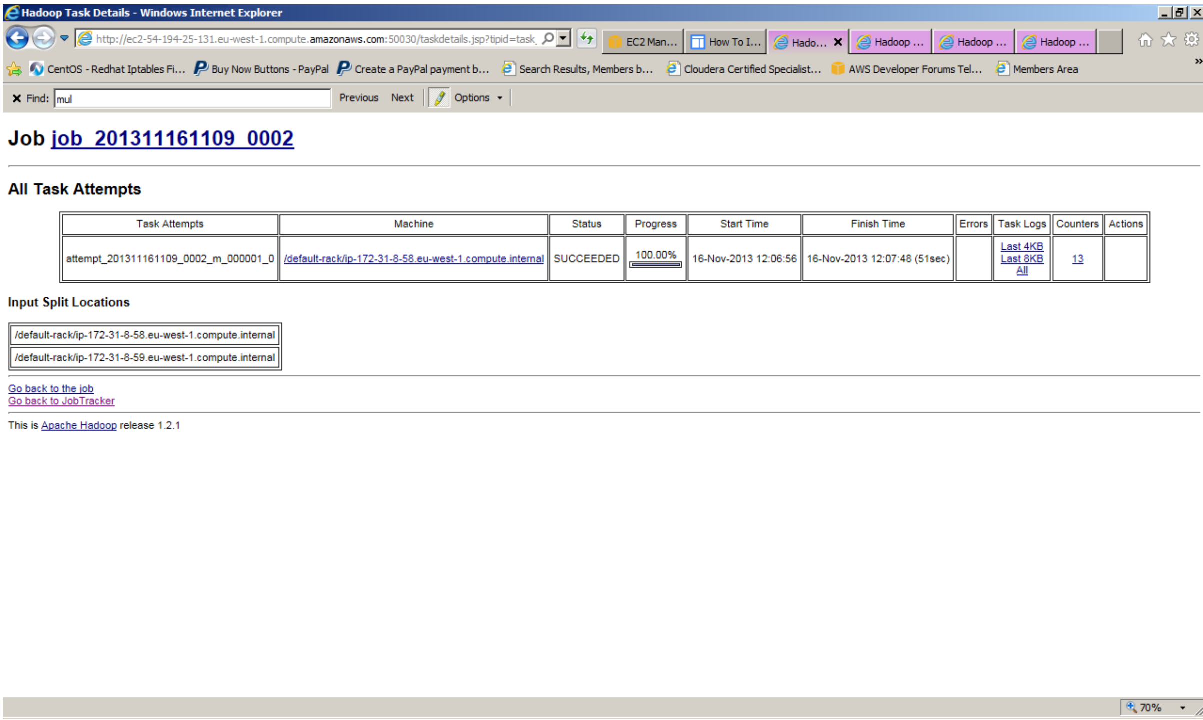
Hadoop job_201311161109_0002 on m - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/jobdetails.jsp?jobid=job_2
EC2 Management Console
How To Install Hortonworks...
Hadoop job_201311161...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Find: mul
Previous
Next
Options
70%

Finished in: 1mins, 10sec
Job Cleanup: Successful

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	4	0	0	4	0	0 / 1
reduce	100.00%	0	0	0	0	0	0 / 0

	Counter	Map	Reduce	Total
Job Counters	SLOTS_MILLIS_MAPS	0	0	238,047
	Total time spent by all reduces waiting after reserving slots (ms)	0	0	0
	Total time spent by all maps waiting after reserving slots (ms)	0	0	0
	Launched map tasks	0	0	5
	Data-local map tasks	0	0	5
	SLOTS_MILLIS_REDUCES	0	0	0
File Output Format Counters	Bytes Written	0	0	219,177,260
File Input Format Counters	Bytes Read	0	0	107,605,652
FileSystemCounters	HDFS_BYTES_READ	107,606,107	0	107,606,107
	FILE_BYTES_WRITTEN	218,444	0	218,444
	HDFS_BYTES_WRITTEN	219,177,260	0	219,177,260
Map-Reduce Framework	Map input records	0	0	2,015,776
	Physical memory (bytes) snapshot	0	0	463,781,888
	Spilled Records	0	0	0
	Total committed heap usage (bytes)	0	0	105,119,744
	CPU time spent (ms)	0	0	48,150
	Virtual memory (bytes) snapshot	0	0	3,347,554,304
	SPLIT_RAW_BYTES	455	0	455
	Map output records	0	0	11,002,956

- Various MapTasks and on Which machine they are excecuted.



Hadoop Task Details - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/taskdetails.jsp?tipid=task_
EC2 Man...
How To I...
Hado...
Hadoop ...
Hadoop ...
Hadoop ...
CentOS - Redhat Iptables Fi...
Buy Now Buttons - PayPal
Create a PayPal payment b...
Search Results, Members b...
Cloudera Certified Specialist...
AWS Developer Forums Tel...
Members Area
Find: mul
Previous
Next
Options
70%

Job job_201311161109_0002

All Task Attempts

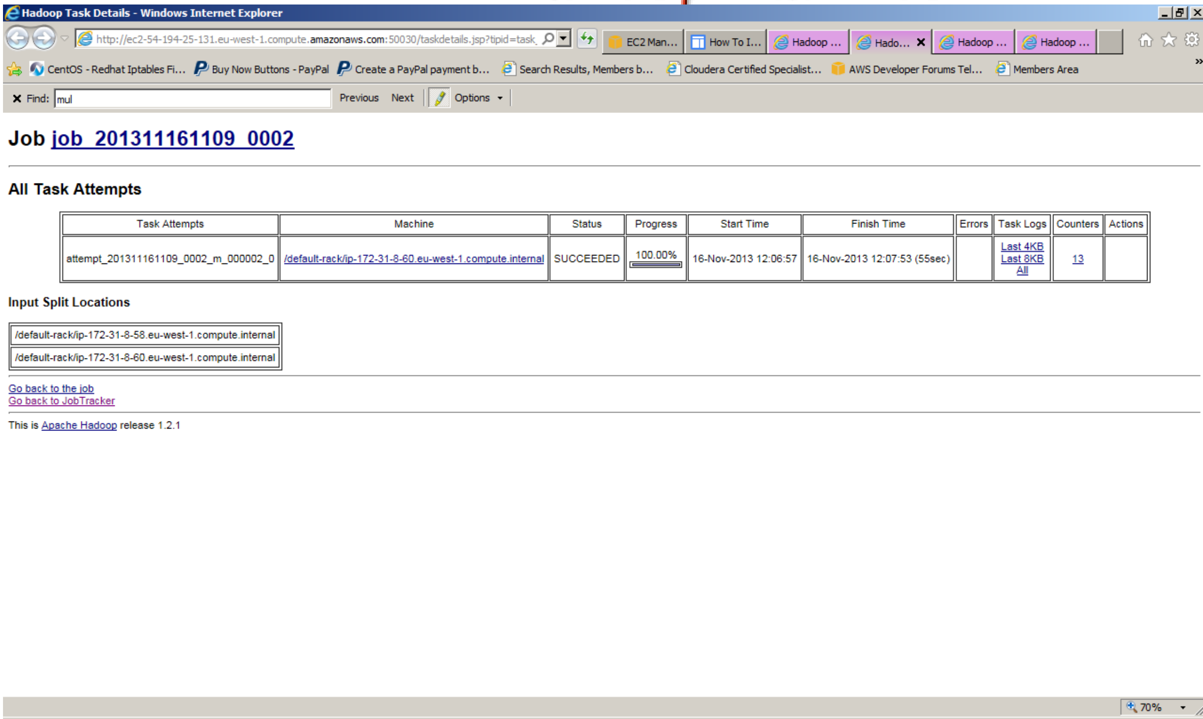
Task Attempts	Machine	Status	Progress	Start Time	Finish Time	Errors	Task Logs	Counters	Actions
attempt_201311161109_0002_m_000001_0	/default-rack/ip-172-31-8-58.eu-west-1.compute.internal	SUCCEEDED	100.00%	16-Nov-2013 12:06:56	16-Nov-2013 12:07:48 (51sec)		Last 4KB Last 8KB All	13	

Input Split Locations

/default-rack/ip-172-31-8-58.eu-west-1.compute.internal
/default-rack/ip-172-31-8-59.eu-west-1.compute.internal

Go back to the job
Go back to JobTracker

This is Apache Hadoop release 1.2.1


Hadoop Task Details - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/taskdetails.jsp?tipid=task_
EC2 Man... How To I... Hadoop ... Hado... Hadoop ... Hadoop ...
CentOS - Redhat Iptables Fi... Buy Now Buttons - PayPal Create a PayPal payment b... Search Results, Members b... Cloudera Certified Specialist... AWS Developer Forums Tel... Members Area
Find: mul Previous Next Options
70%

Job job_201311161109_0002

All Task Attempts

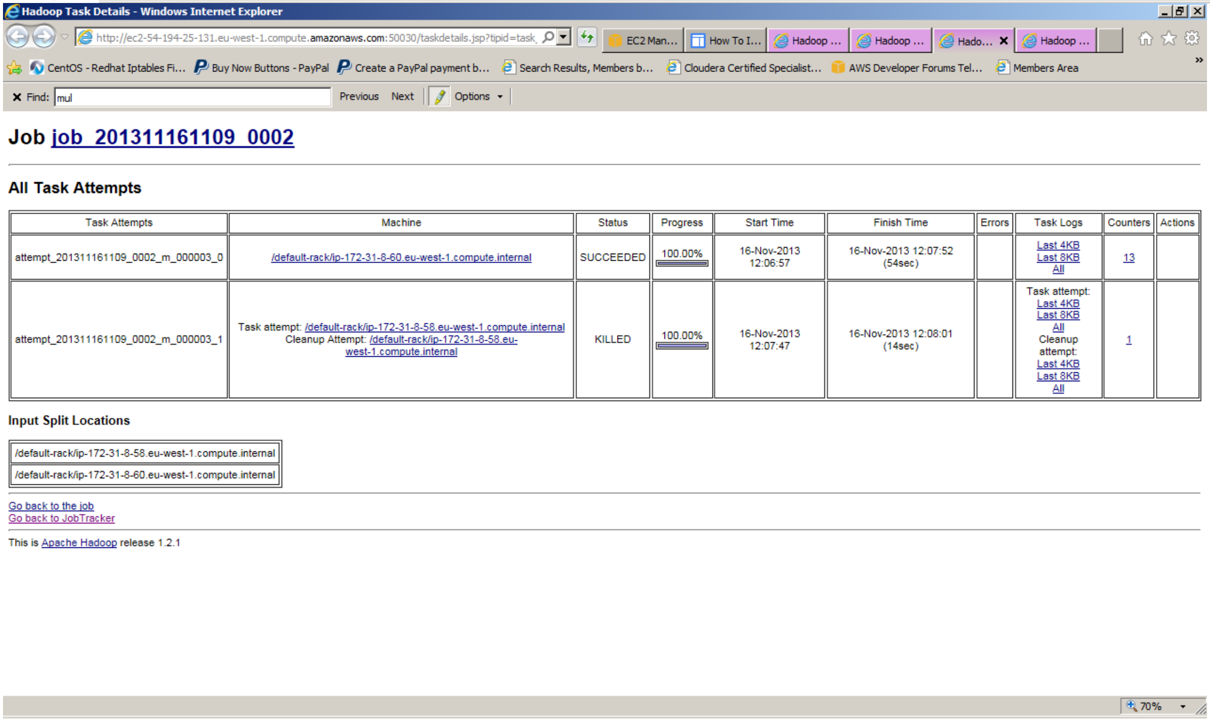
Task Attempts	Machine	Status	Progress	Start Time	Finish Time	Errors	Task Logs	Counters	Actions
attempt_201311161109_0002_m_000002_0	/default-rack/ip-172-31-8-60.eu-west-1.compute.internal	SUCCEEDED	100.00%	16-Nov-2013 12:06:57	16-Nov-2013 12:07:53 (55sec)		Last 4KB Last 8KB All	13	

Input Split Locations

/default-rack/ip-172-31-8-58.eu-west-1.compute.internal
/default-rack/ip-172-31-8-60.eu-west-1.compute.internal

Go back to the job
Go back to JobTracker

This is Apache Hadoop release 1.2.1

- 
Hadoop Task Details - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/taskdetails.jsp?tipid=task_
EC2 Man... How To I... Hadoop ... Hadoop ... Hado... Hadoop ...
CentOS - Redhat Iptables Fi... Buy Now Buttons - PayPal Create a PayPal payment b... Search Results, Members b... Cloudera Certified Specialist... AWS Developer Forums Tel... Members Area
Find: mul Previous Next Options
70%

Job job_201311161109_0002

All Task Attempts

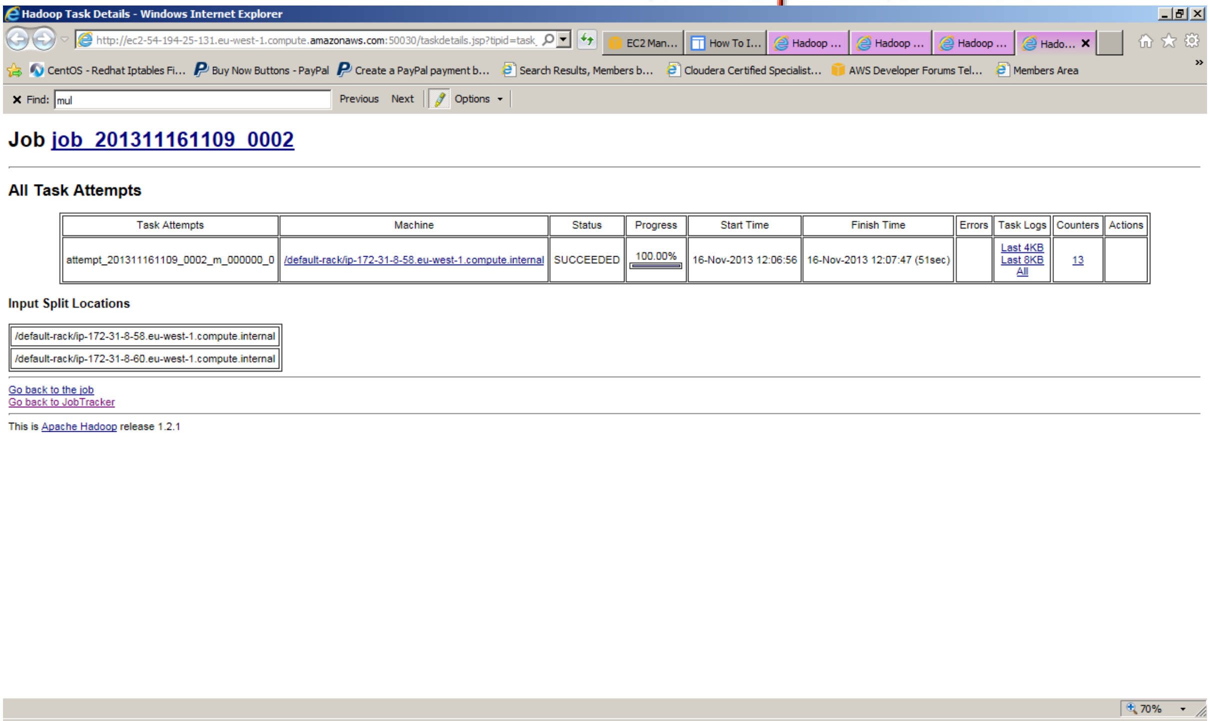
Task Attempts	Machine	Status	Progress	Start Time	Finish Time	Errors	Task Logs	Counters	Actions
attempt_201311161109_0002_m_000003_0	/default-rack/ip-172-31-8-60.eu-west-1.compute.internal	SUCCEEDED	100.00%	16-Nov-2013 12:06:57	16-Nov-2013 12:07:52 (54sec)		Last 4KB Last 8KB All	13	
attempt_201311161109_0002_m_000003_1	Task attempt: /default-rack/ip-172-31-8-58.eu-west-1.compute.internal Cleanup Attempt: /default-rack/ip-172-31-8-58.eu-west-1.compute.internal	KILLED	100.00%	16-Nov-2013 12:07:47	16-Nov-2013 12:08:01 (14sec)		Task attempt: Last 4KB Last 8KB All Cleanup attempt: Last 4KB Last 8KB All	1	

Input Split Locations

/default-rack/ip-172-31-8-58.eu-west-1.compute.internal
/default-rack/ip-172-31-8-60.eu-west-1.compute.internal

Go back to the job
Go back to JobTracker

This is Apache Hadoop release 1.2.1



Hadoop Task Details - Windows Internet Explorer
http://ec2-54-194-25-131.eu-west-1.compute.amazonaws.com:50030/taskdetails.jsp?tipid=task_
Find: mul
Previous
Next
Options
70%

Job job_201311161109_0002

All Task Attempts

Task Attempts	Machine	Status	Progress	Start Time	Finish Time	Errors	Task Logs	Counters	Actions
attempt_201311161109_0002_m_000000_0	/default-rack/ip-172-31-8-58.eu-west-1.compute.internal	SUCCEEDED	100.00%	16-Nov-2013 12:06:56	16-Nov-2013 12:07:47 (51sec)		Last 4KB Last 8KB All	13	

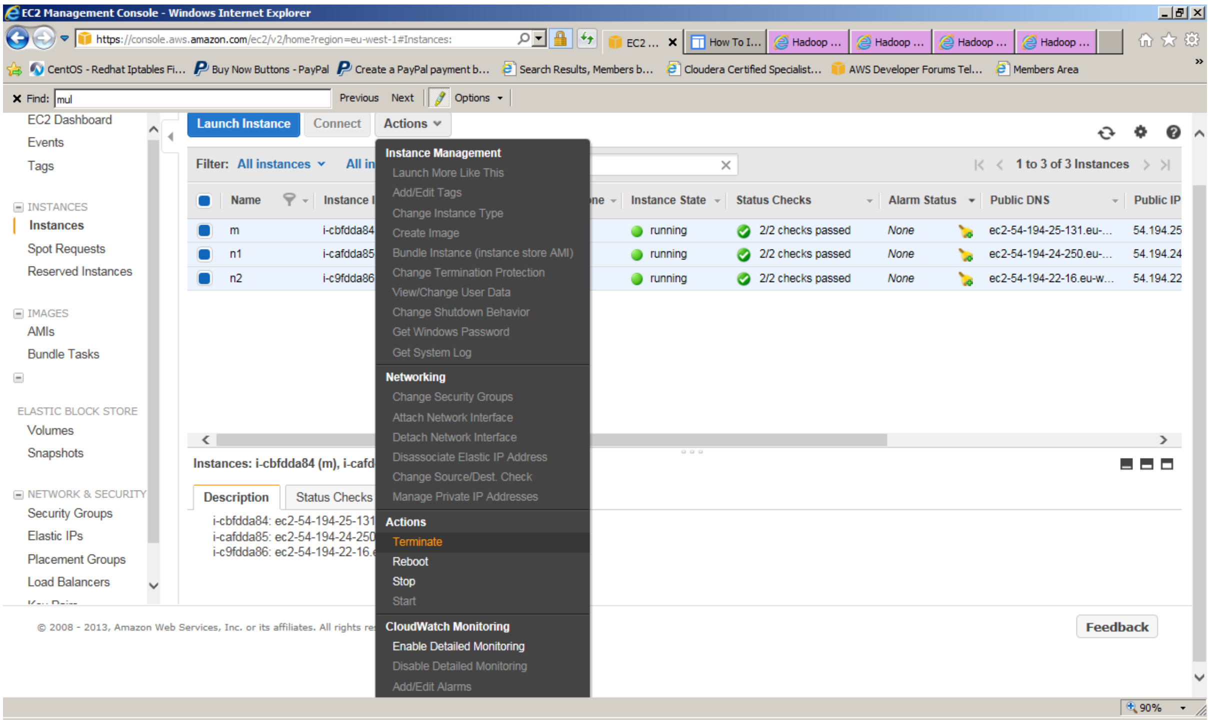
Input Split Locations

/default-rack/ip-172-31-8-58.eu-west-1.compute.internal
/default-rack/ip-172-31-8-60.eu-west-1.compute.internal

Go back to the job
Go back to JobTracker

This is Apache Hadoop release 1.2.1

41. Terminate the all Instances to avoid recurring costs as below.

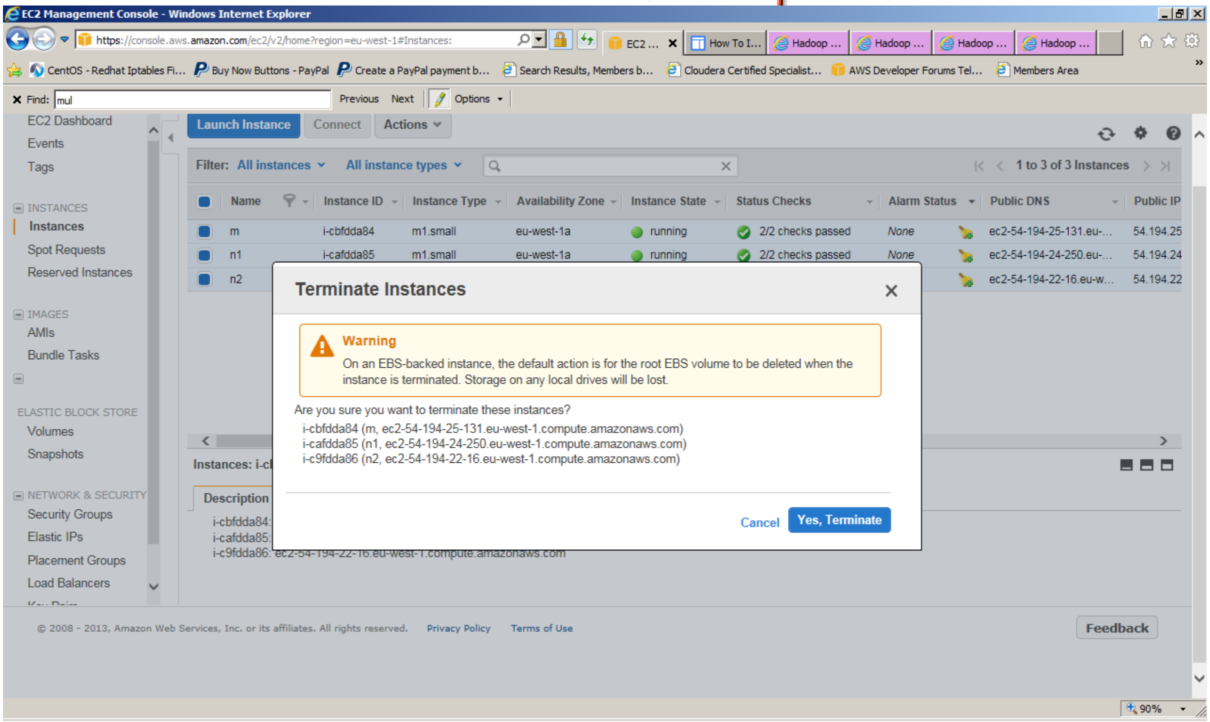


Name	Instance I	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25
n1	i-cafdda85	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24
n2	i-c9fdda86	running	2/2 checks passed	None	ec2-54-194-22-16.eu-w...	54.194.22

Instances: i-cbfdda84 (m), i-cafd

i-cbfdda84: ec2-54-194-25-131
i-cafdda85: ec2-54-194-24-250
i-c9fdda86: ec2-54-194-22-16.

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights re



Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP
m	i-cbfdda84	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-25-131.eu-...	54.194.25
n1	i-cafdda85	m1.small	eu-west-1a	running	2/2 checks passed	None	ec2-54-194-24-250.eu-...	54.194.24
n2							ec2-54-194-22-16.eu-w...	54.194.22

Terminate Instances

Warning

On an EBS-backed instance, the default action is for the root EBS volume to be deleted when the instance is terminated. Storage on any local drives will be lost.

Are you sure you want to terminate these instances?

i-cbfdda84 (m, ec2-54-194-25-131.eu-west-1.compute.amazonaws.com)
i-cafdda85 (n1, ec2-54-194-24-250.eu-west-1.compute.amazonaws.com)
i-c9fdda86 (n2, ec2-54-194-22-16.eu-west-1.compute.amazonaws.com)

Cancel Yes, Terminate

© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

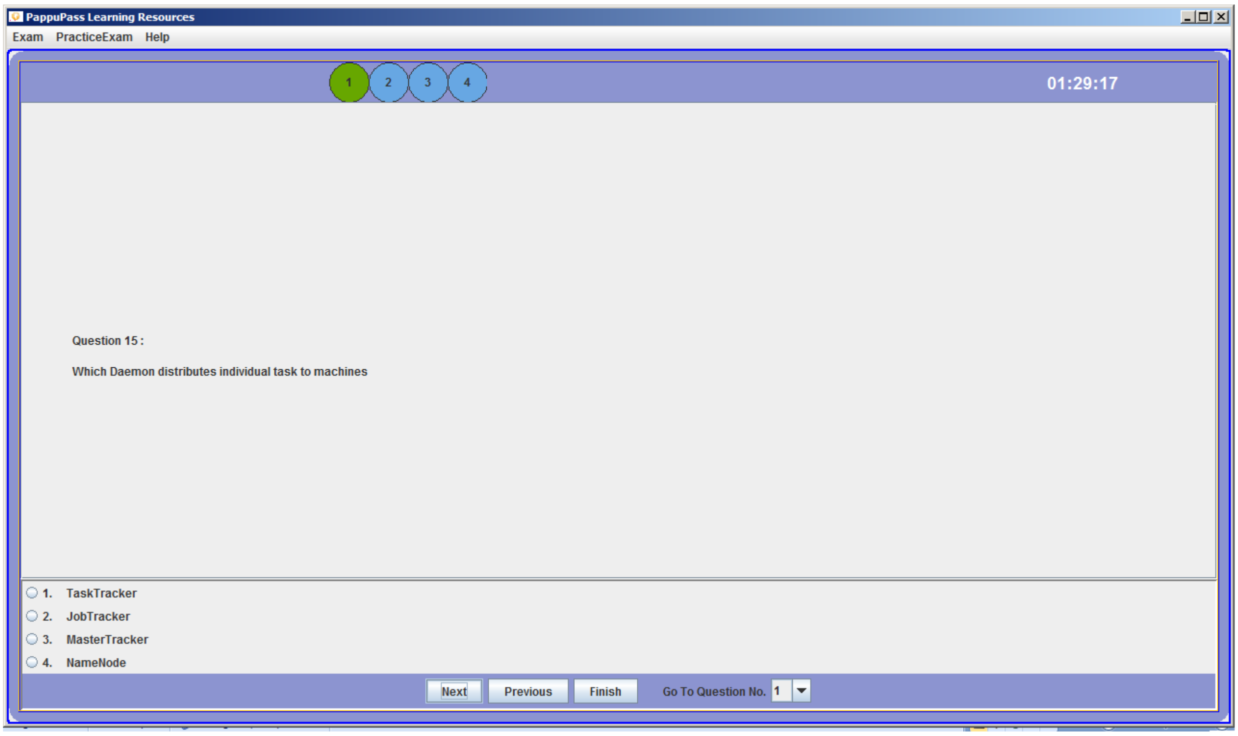
42. Happy Hadoop Learning....
43. Send your suggestions to us admin@hadoopexam.com

HadoopExam.com

Hadoop Certification Exam Simulator (Developer/Administrator/HBase) + Study Material

- Contains 4 practice Question Paper
- 240/238/214 (Developer/Admin/HBase) realistic Hadoop Certification Questions
- All Questions are on latest Pattern
- End time 15 Page revision notes for Developer (Save lot of time)
- Download from www.HadoopExam.com

Note: There is 50% talent gap in BigData domain, get Hadoop certification with the HadoopExam Learning Resources Hadoop Exam Simulator.



Question 15 :

Which Daemon distributes individual task to machines

○ 1. TaskTracker
○ 2. JobTracker
○ 3. MasterTracker
○ 4. NameNode