HADOOP CLUSTER SETUP ON GOOGLE CLOUD COMPUTE

Instruction Manual

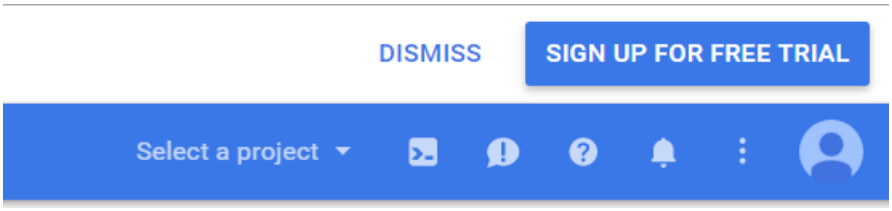
FEBRUARY 5, 2016

PRABHU MOORTHY
prabhu88@hotmail.com

Contents

Creating a Google Cloud Account

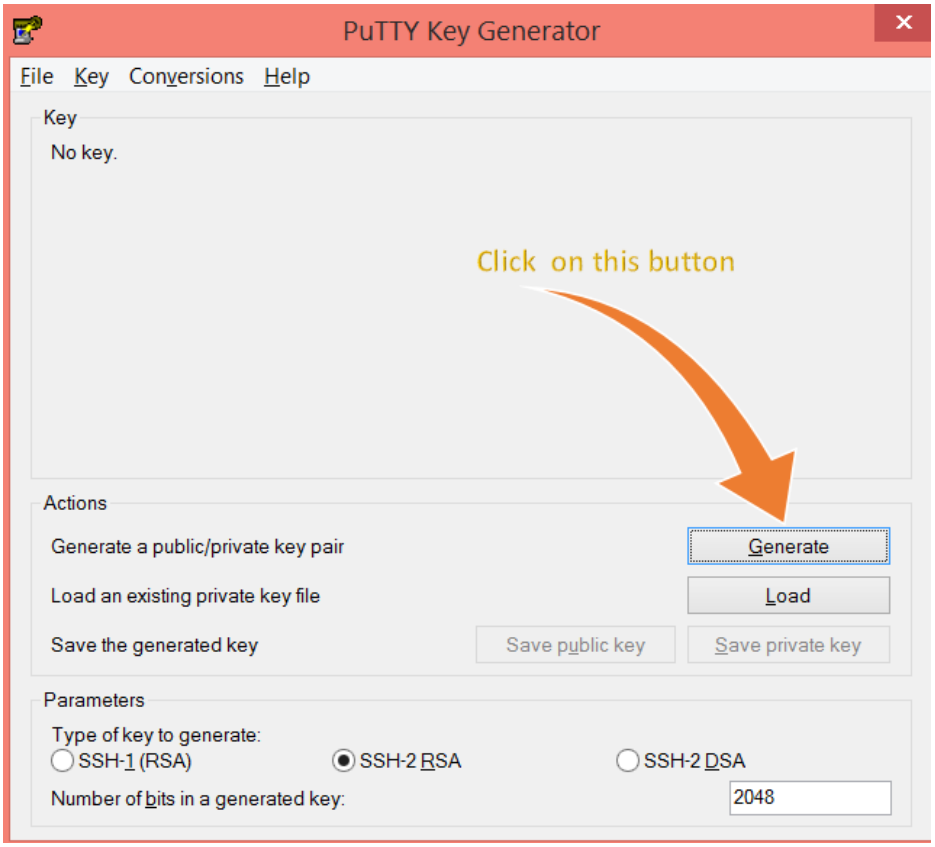
1) Go to https://cloud.google.com/compute/
2) Click on the **Sign in** button.
3) Sign in with your Gmail id and password or create a new one.
4) Once logged in click on My Console.
5) On the next page click on **SIGN UP FOR FREE TRIAL** for a free trial.
6) **On the free trial you'll get $300 in credit and 60 days to explore Google Cloud Platform.**

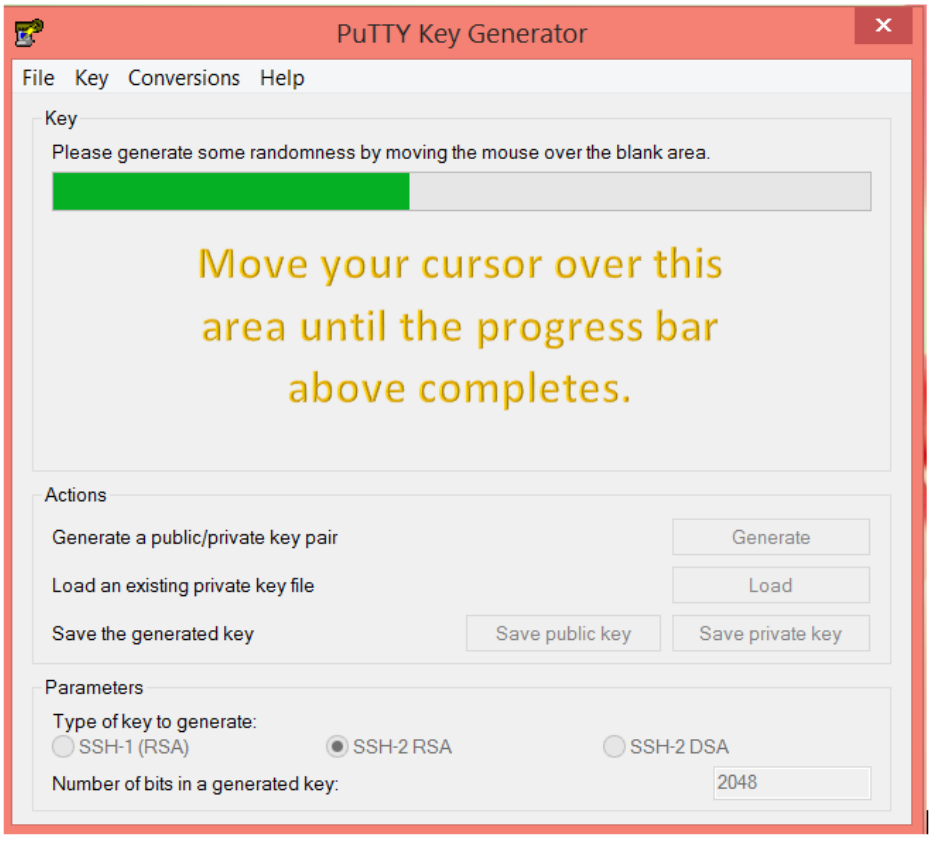


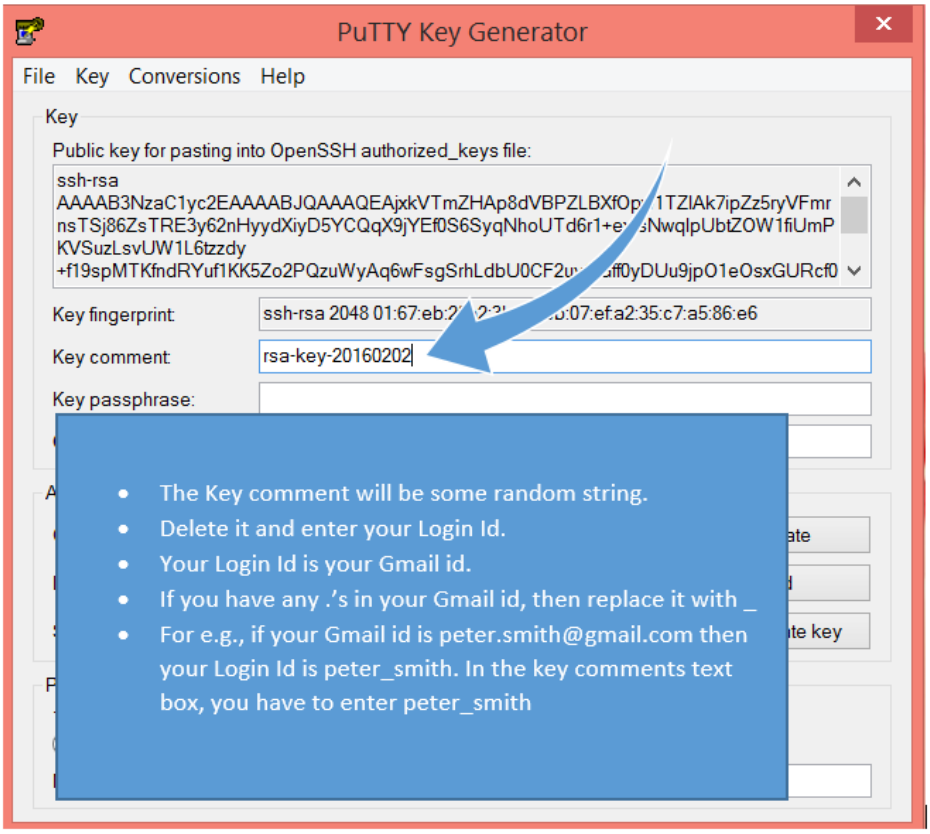
7) Fill out the form on the next page, you will have to give your credit card details. Do not worry, only 1$ will be charged on your credit card to verify.
8) Remember to deactivate your Google Cloud Account on the end of 60 days to avoid any charges. Google does not take any amount from your card without your permission, even if you forget to deactivate after 60 days but to be on a safer side it is advisable to deactivate the account.

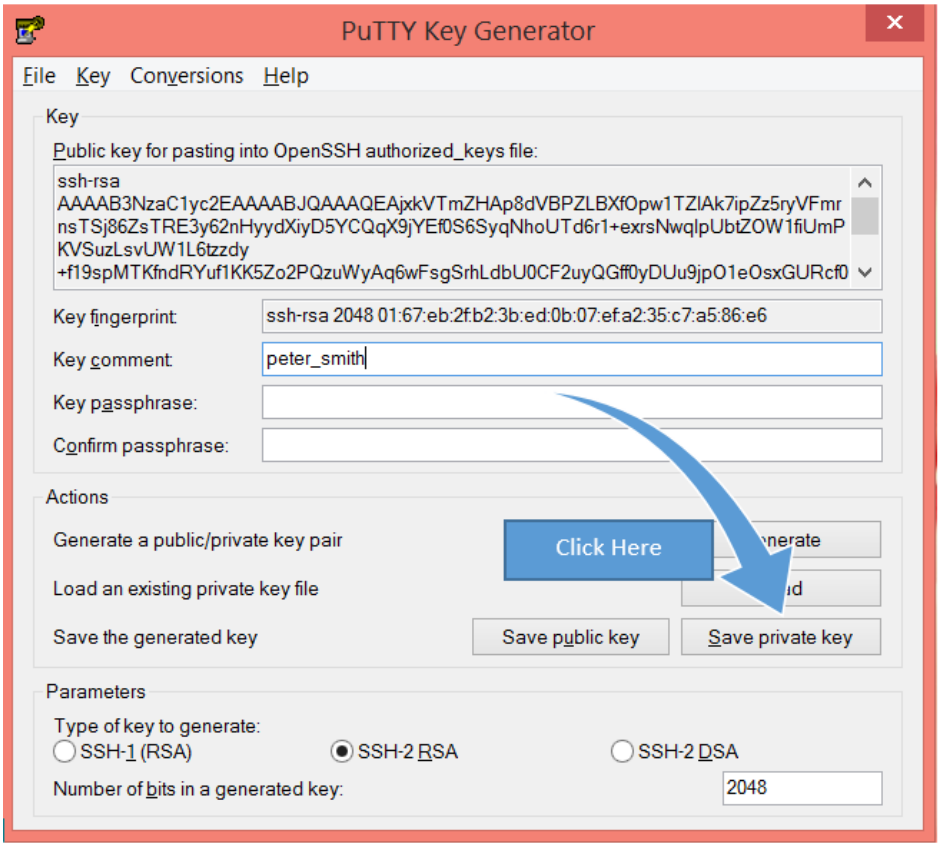
Generating Public/Private key

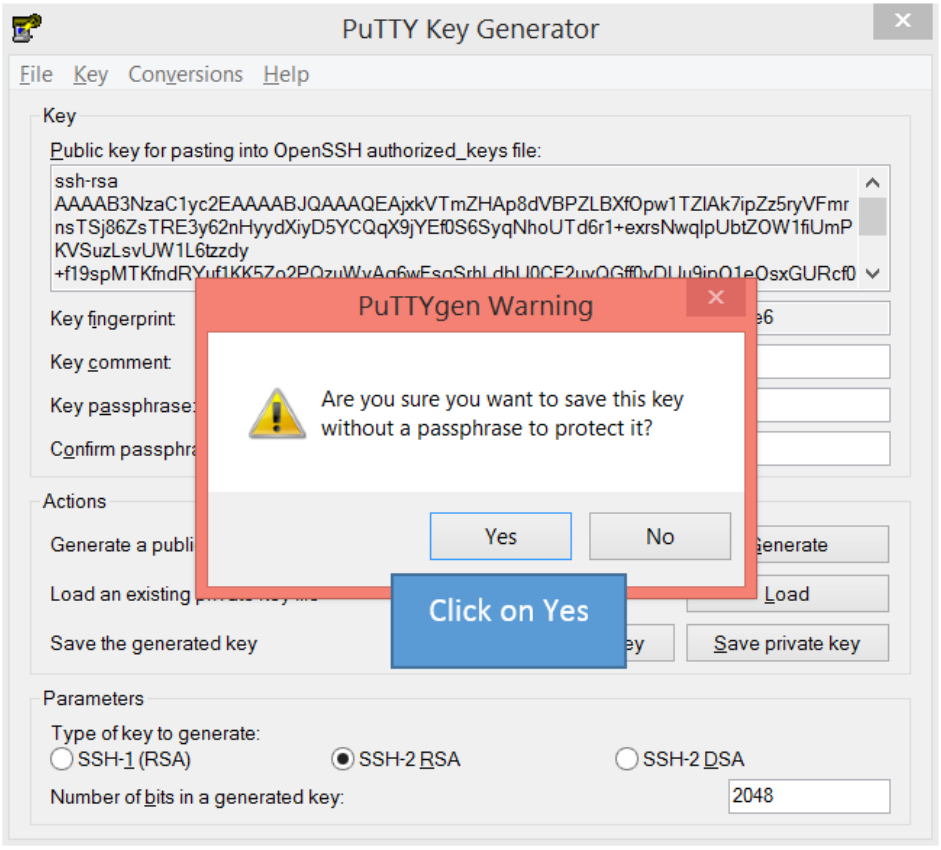
1) Download puttygen application.
2) Open puttygen application and follow the below steps.

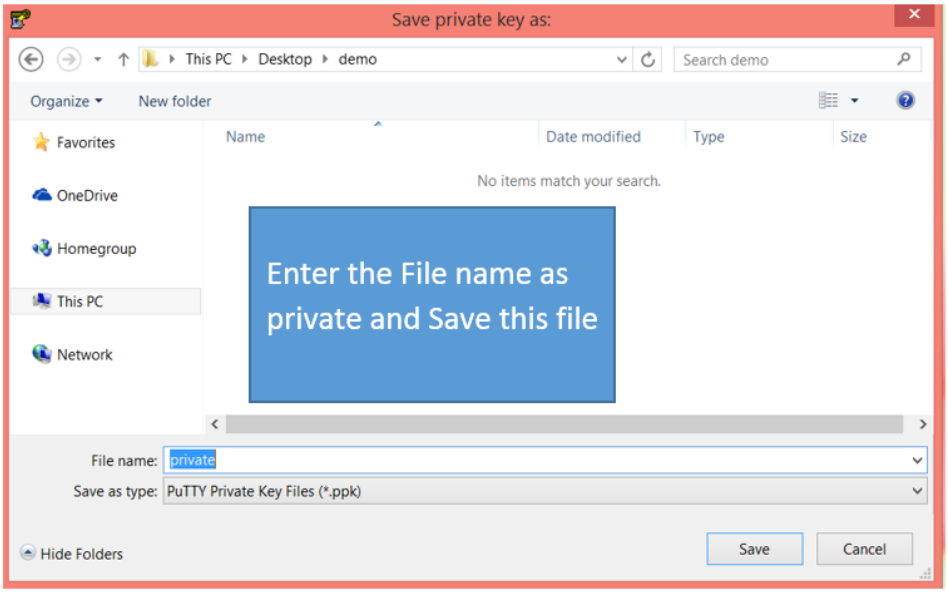


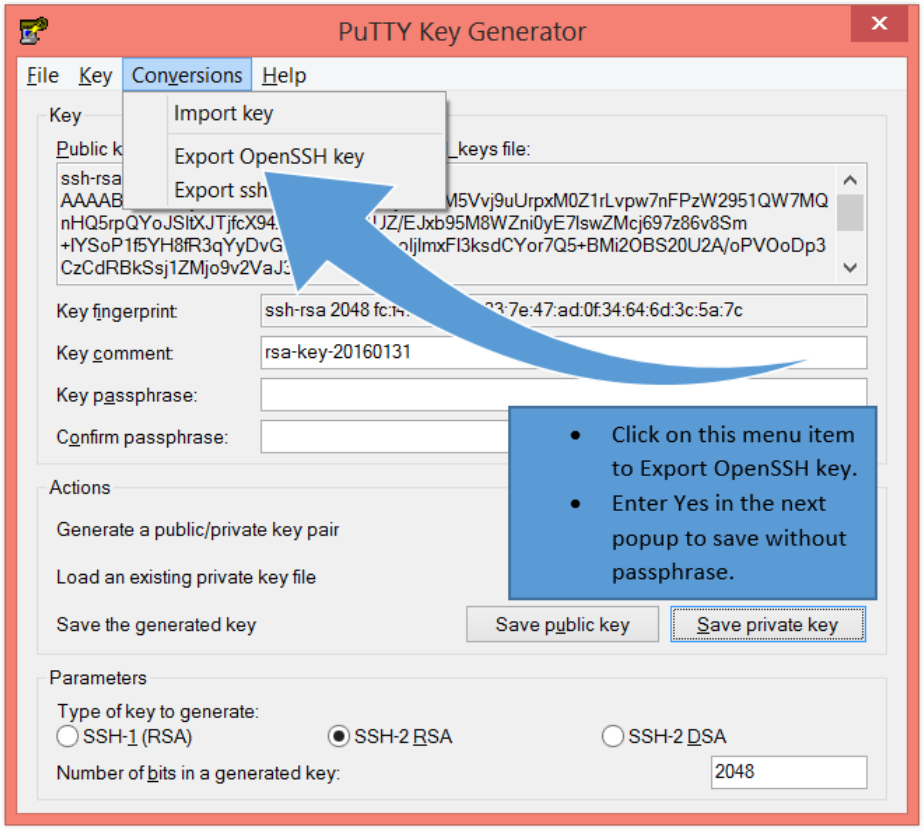


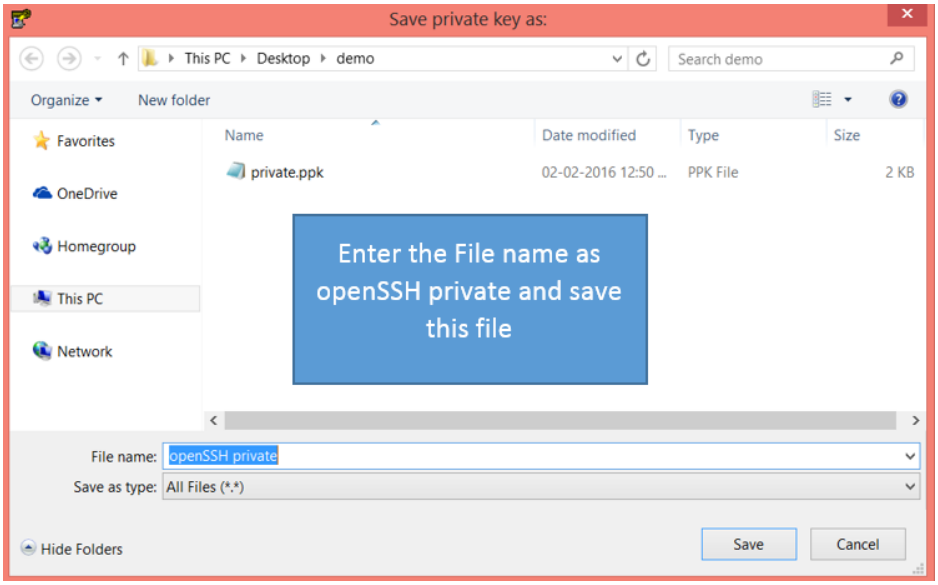

PuTTY Key Generator
File Key Conversions Help
Key
Public key for pasting into OpenSSH authorized_keys file:
ssh-rsa
AAAAB3NzaC1yc2EAAAABJQAAAQEAjxkVTmZHAp8dVBPZLBXfOpw1TZlAk7ipZz5ryVFmr
nsTSj86ZsTRE3y62nHyydXiyD5YCQqX9jYEf0S6SyqNhoUTd6r1+exrsNwqlpUbtZOW1fiUmP
KVSuzLsvUW1L6tzzdy
+f19spMTKfndRYuf1KK5Zo2PQzuWyAq6wFsgSrhLdbU0CF2uyQGff0yDUu9jpO1eOsxGURcf0
Key fingerprint:
ssh-rsa 2048 01:67:eb:2f:b2:3b:ed:0b:07:ef:a2:35:c7:a5:86:e6
Key comment:
rsa-key-20160202
Key passphrase:
The Key comment will be some random string.
Delete it and enter your Login Id.
Your Login Id is your Gmail id.
If you have any .'s in your Gmail id, then replace it with _
For e.g., if your Gmail id is peter.smith@gmail.com then your Login Id is peter_smith. In the key comments text box, you have to enter peter_smith

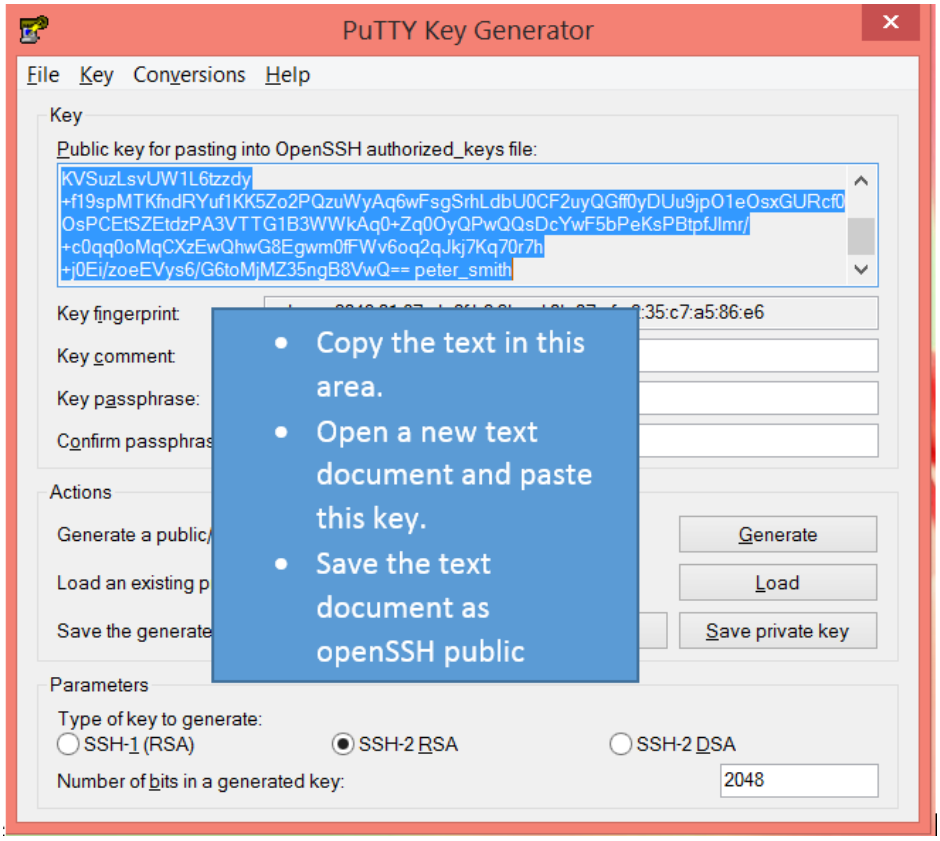

PuTTY Key Generator
File Key Conversions Help
Key
Public key for pasting into OpenSSH authorized_keys file:
ssh-rsa
AAAAB3NzaC1yc2EAAAABJQAAAQEAjxkVTmZHAp8dVBPZLBXfOpw1TZlAk7ipZz5ryVFmr
nsTSj86ZsTRE3y62nHyydXiyD5YCQqX9jYEf0S6SyqNhoUTd6r1+exrsNwqlpUbtZOW1fiUmP
KVSuzLsvUW1L6tzzdy
+f19spMTKfndRYuf1KK5Zo2PQzuWyAq6wFsgSrhLdbU0CF2uyQGff0yDUu9jpO1eOsxGURcf0
Key fingerprint:
ssh-rsa 2048 01:67:eb:2f:b2:3b:ed:0b:07:ef:a2:35:c7:a5:86:e6
Key comment:
peter_smith
Key passphrase:
Confirm passphrase:
Actions
Generate a public/private key pair
Click Here
Load an existing private key file
Save the generated key
Save public key
Save private key
Parameters
Type of key to generate:
SSH-1 (RSA)
SSH-2 RSA
SSH-2 DSA
Number of bits in a generated key:
2048


PuTTY Key Generator
File Key Conversions Help
Key
Public key for pasting into OpenSSH authorized_keys file:
ssh-rsa
AAAAB3NzaC1yc2EAAAABJQAAAQEAjxkVTmZHAp8dVBPZLBXfOpw1TZlAk7ipZz5ryVFmr
nsTSj86ZsTRE3y62nHyydXiyD5YCQqX9jYEf0S6SyqNhoUTd6r1+exrsNwqlpUbtZOW1fiUmP
KVSuzLsvUW1L6tzzdy
Key fingerprint:
Key comment:
Key passphrase:
Confirm passphra
Actions
Generate a publi
Load an existing
Save the generated key
Load
Save private key
PuTTYgen Warning
Are you sure you want to save this key without a passphrase to protect it?
Yes
No
Click on Yes
Parameters
Type of key to generate:
SSH-1 (RSA)
SSH-2 RSA
SSH-2 DSA
Number of bits in a generated key:
2048


Save private key as:
This PC ▸ Desktop ▸ demo
Search demo
Organize
New folder
Favorites
OneDrive
Homegroup
This PC
Network
Name
Date modified
Type
Size
No items match your search.
Enter the File name as private and Save this file
File name: private
Save as type: PuTTY Private Key Files (*.ppk)
Hide Folders
Save
Cancel


PuTTY Key Generator
File Key Conversions Help
Import key
Export OpenSSH key
Export ssh
Key
Public k
keys file:
ssh-rsa
AAAAB
M5Vvj9uUrpxM0Z1rLvpw7nFPzW2951QW7MQ
nHQ5rpQYoJSltXJTjfcX94
UZ/EJxb95M8WZni0yE7lswZMcj697z86v8Sm
+lYSoP1f5YH8fR3qYyDvG
oljImxFl3ksdCYor7Q5+BMi2OBS20U2A/oPVOoDp3
CzCdRBkSsj1ZMjo9v2VaJ
Key fingerprint:
ssh-rsa 2048 fc:
:7e:47:ad:0f:34:64:6d:3c:5a:7c
Key comment:
rsa-key-20160131
Key passphrase:
Confirm passphrase:
Actions
Generate a public/private key pair
Load an existing private key file
Save the generated key
Save public key
Save private key
Click on this menu item to Export OpenSSH key.
Enter Yes in the next popup to save without passphrase.
Parameters
Type of key to generate:
SSH-1 (RSA)
SSH-2 RSA
SSH-2 DSA
Number of bits in a generated key:
2048

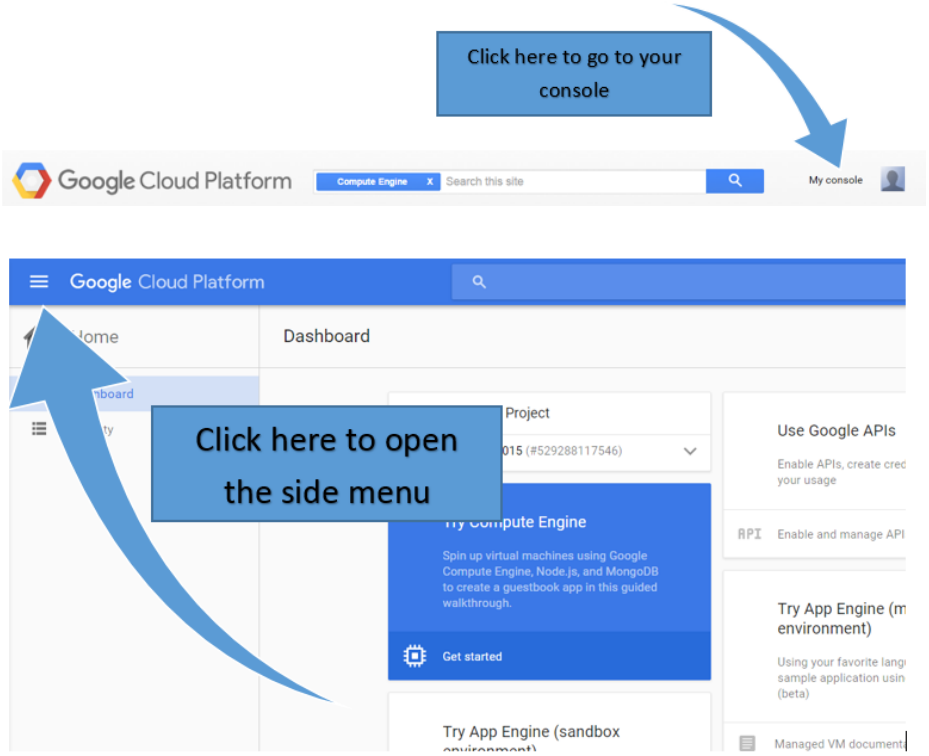

Save private key as:
This PC ▸ Desktop ▸ demo
Search demo
Organize
New folder
Favorites
OneDrive
Homegroup
This PC
Network
Name
Date modified
Type
Size
private.ppk
02-02-2016 12:50 ...
PPK File
2 KB
Enter the File name as openSSH private and save this file
File name:
openSSH private
Save as type:
All Files (*.*)
Hide Folders
Save
Cancel

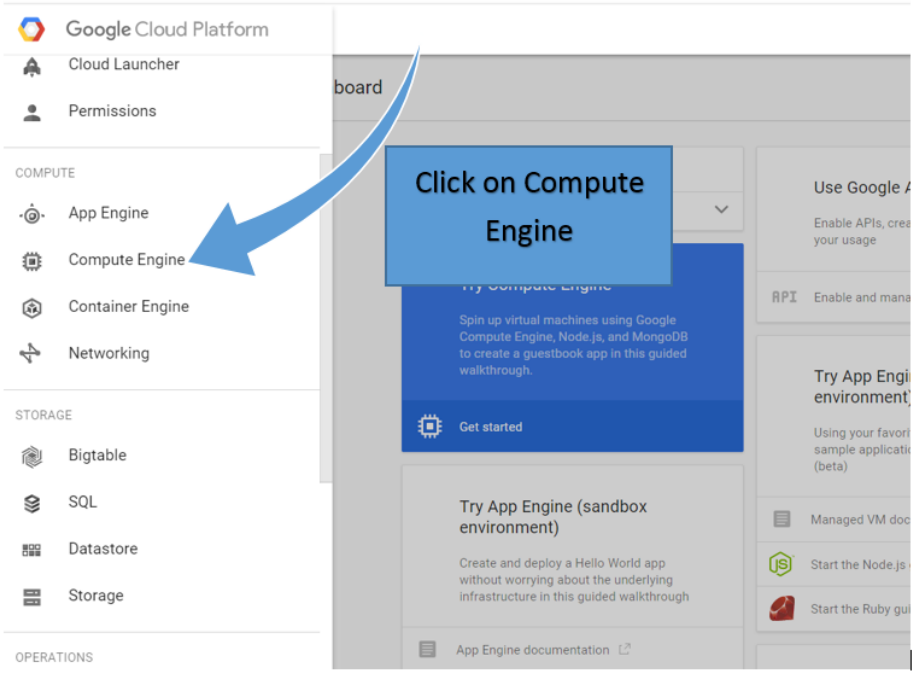


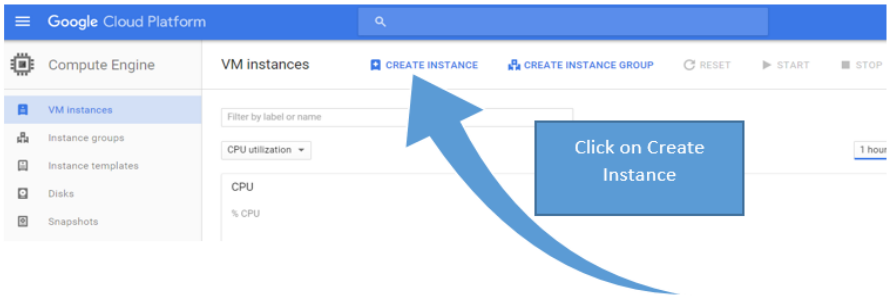
3) Now you should be having 3 keys saved in your local desktop.
    1. private.ppk
    2. openSSH private
    3. openSSH public.txt
4) Keep these keys saved somewhere safe, you will be needing these keys for as long as you have your cluster in the cloud.
5) Close puttygen application.

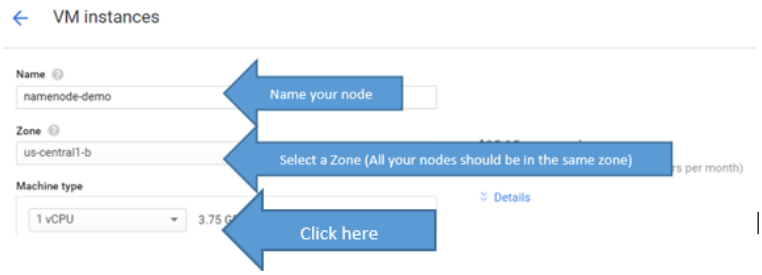
Creating a Cluster

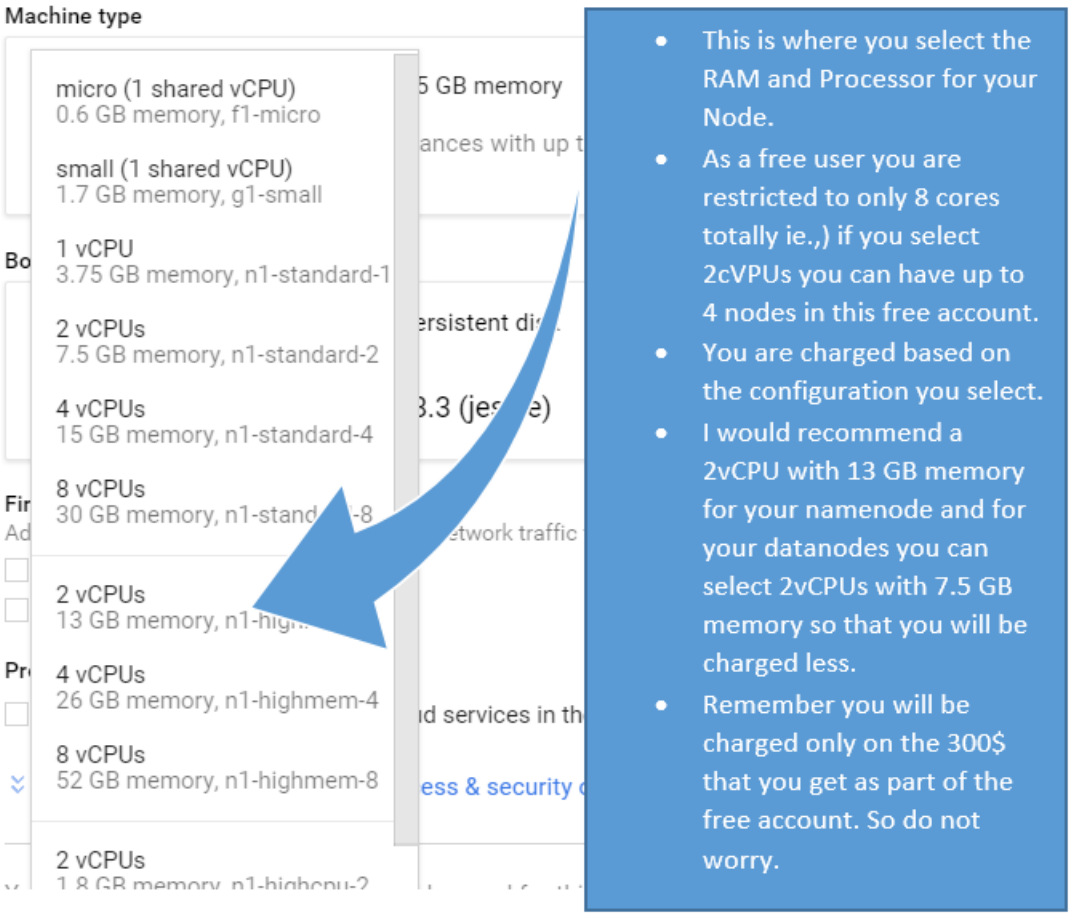
1) Now, go to the Google cloud compute web page and follow the below steps.



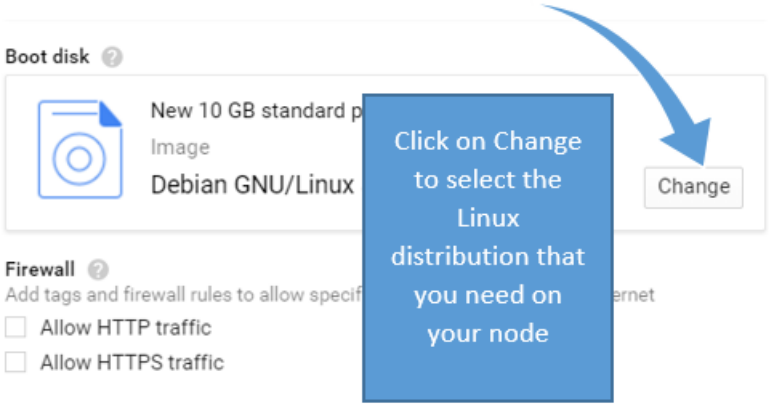


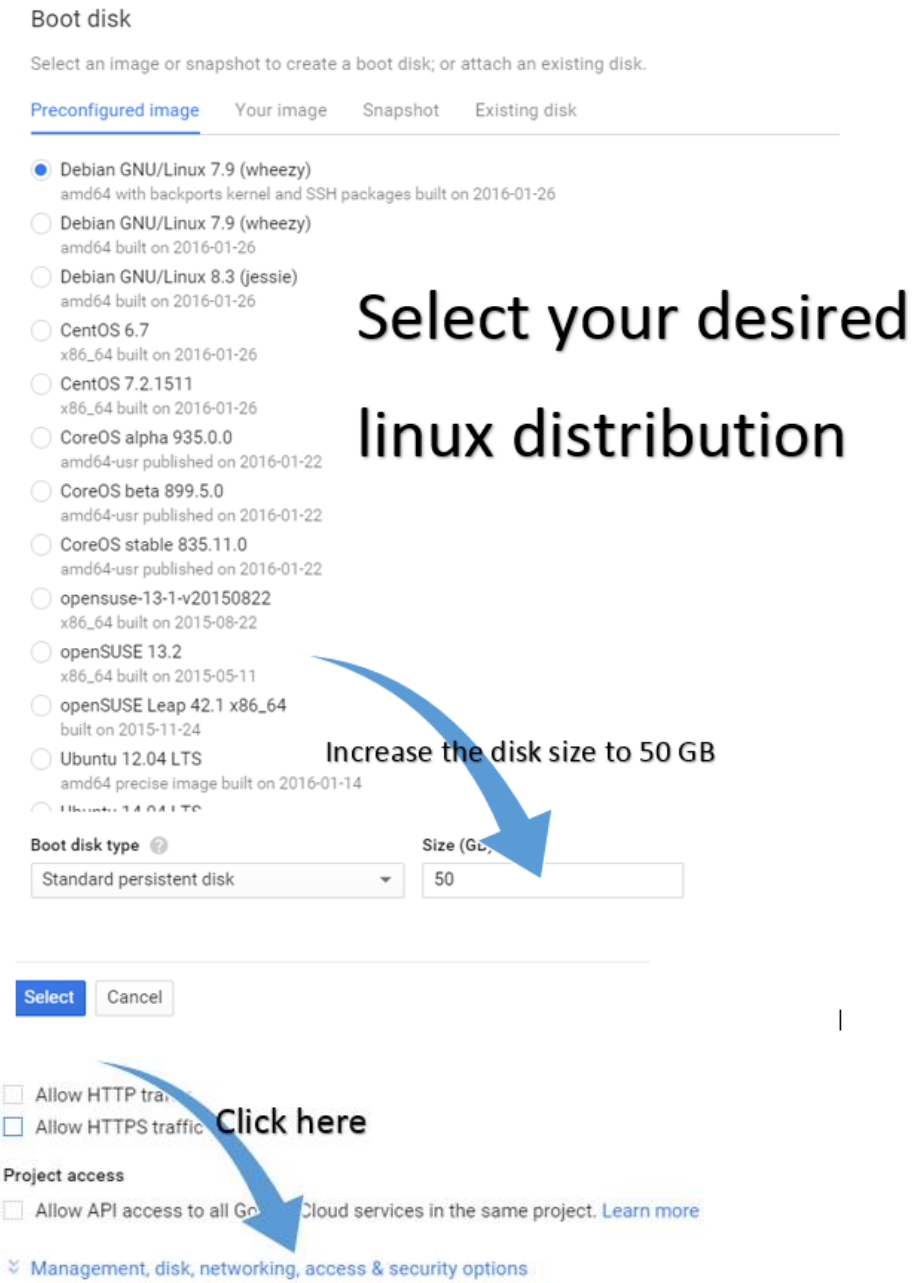




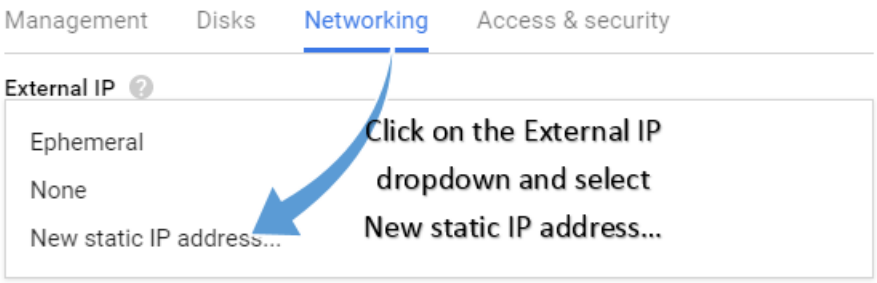


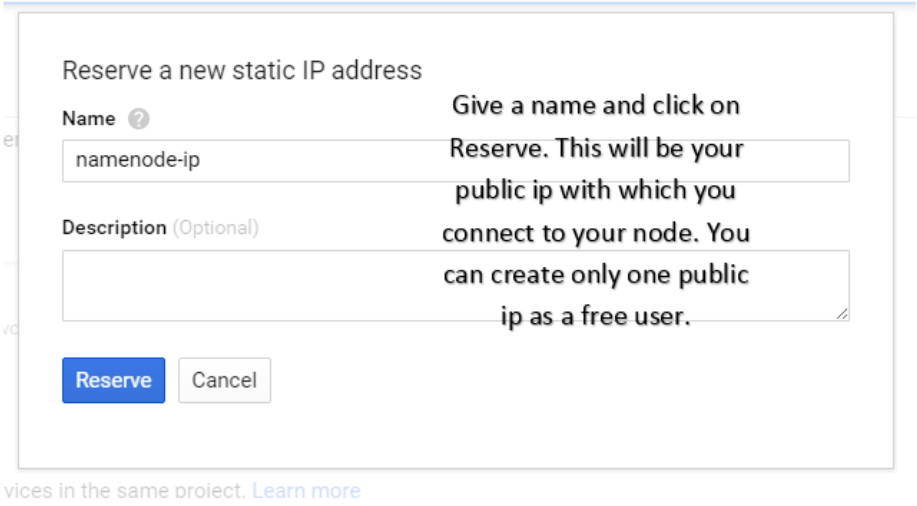
- This is where you select the RAM and Processor for your Node.
- As a free user you are restricted to only 8 cores totally ie.,) if you select 2cVPUs you can have up to 4 nodes in this free account.
- You are charged based on the configuration you select.
- I would recommend a 2vCPU with 13 GB memory for your namenode and for your datanodes you can select 2vCPUs with 7.5 GB memory so that you will be charged less.
- Remember you will be charged only on the 300$ that you get as part of the free account. So do not worry.

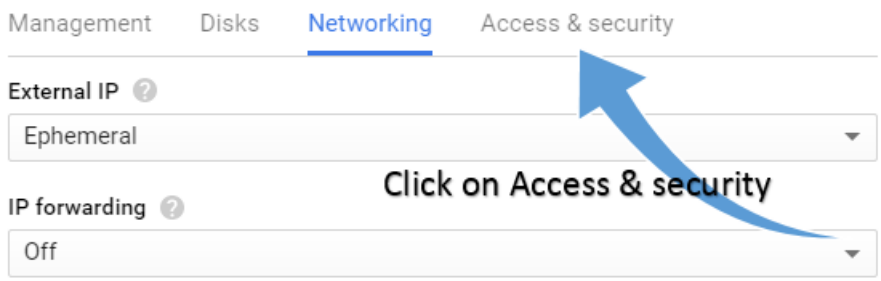

Boot disk
New 10 GB standard p
Image
Debian GNU/Linux
Change
Click on Change to select the Linux distribution that you need on your node
Firewall
Add tags and firewall rules to allow specif
ernet
Allow HTTP traffic
Allow HTTPS traffic

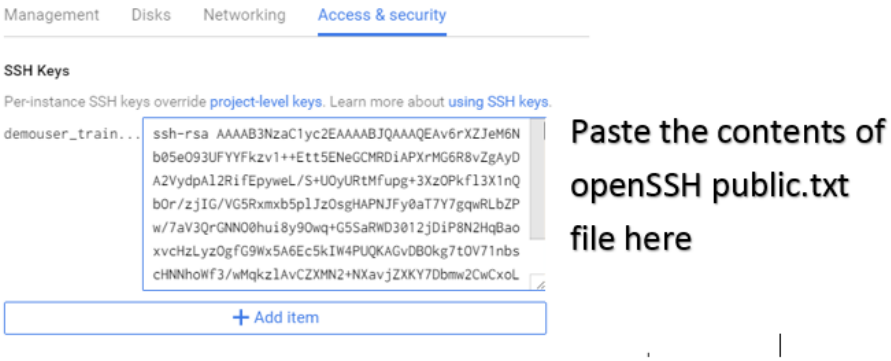

Boot disk
Select an image or snapshot to create a boot disk; or attach an existing disk.
Preconfigured image
Your image
Snapshot
Existing disk
Debian GNU/Linux 7.9 (wheezy)
amd64 with backports kernel and SSH packages built on 2016-01-26
Debian GNU/Linux 7.9 (wheezy)
amd64 built on 2016-01-26
Debian GNU/Linux 8.3 (jessie)
amd64 built on 2016-01-26
CentOS 6.7
x86_64 built on 2016-01-26
CentOS 7.2.1511
x86_64 built on 2016-01-26
CoreOS alpha 935.0.0
amd64-usr published on 2016-01-22
CoreOS beta 899.5.0
amd64-usr published on 2016-01-22
CoreOS stable 835.11.0
amd64-usr published on 2016-01-22
opensuse-13-1-v20150822
x86_64 built on 2015-08-22
openSUSE 13.2
x86_64 built on 2015-05-11
openSUSE Leap 42.1 x86_64
built on 2015-11-24
Ubuntu 12.04 LTS
amd64 precise image built on 2016-01-14
Select your desired linux distribution
Increase the disk size to 50 GB
Boot disk type
Standard persistent disk
50
Select
Cancel
Allow HTTP tra
Allow HTTPS traffic
Click here
Project access
Allow API access to all Go
Cloud services in the same project. Learn more
Management, disk, networking, access & security options

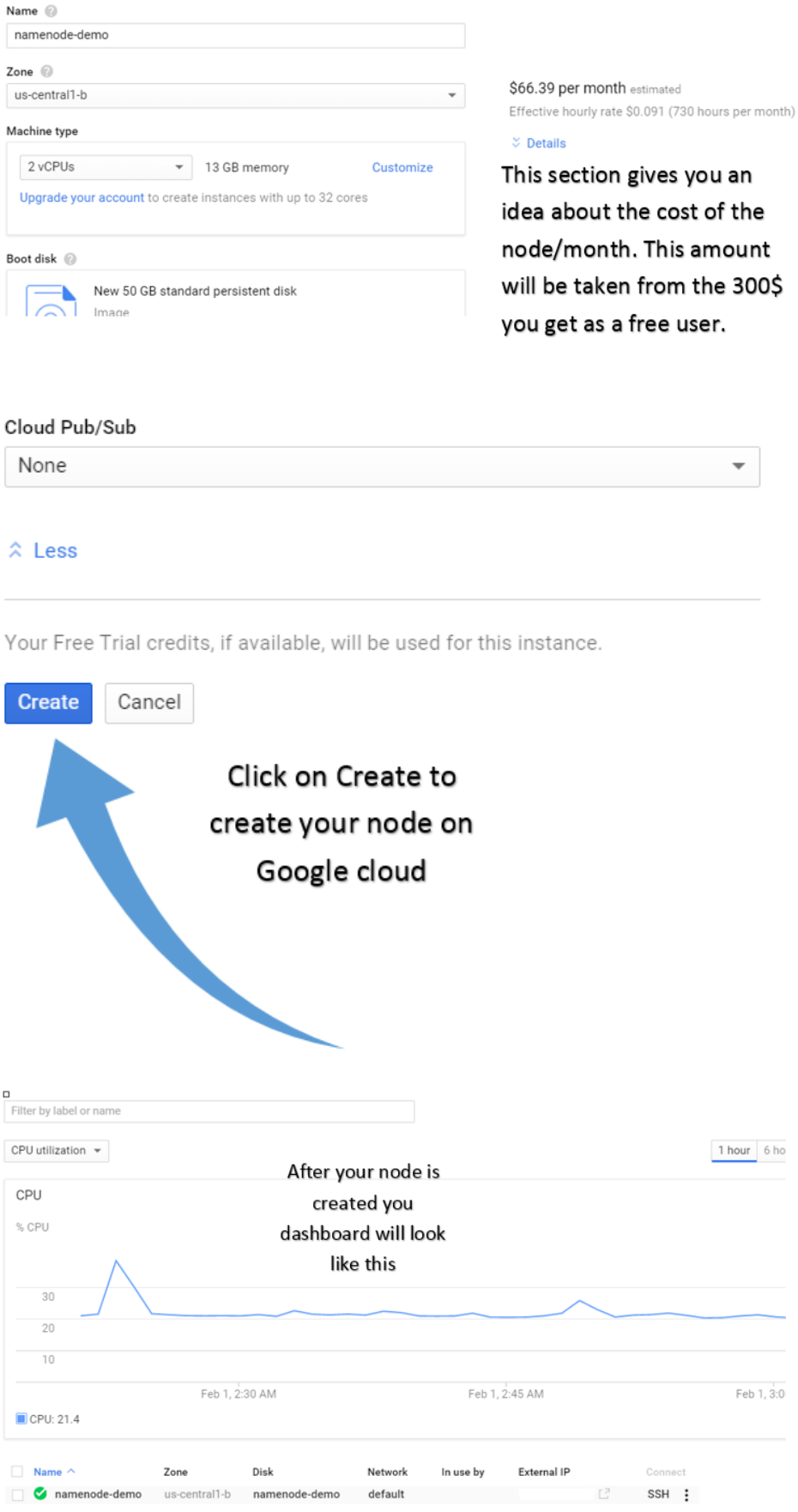

Management
Disks
Networking
Access & security
Description (Optional)
Click on Networking
Tags (Optional)


Management
Disks
Networking
Access & security
External IP
Ephemeral
None
New static IP address...
Click on the External IP dropdown and select New static IP address...


Reserve a new static IP address
Name
namenode-ip
Description (Optional)
Give a name and click on Reserve. This will be your public ip with which you connect to your node. You can create only one public ip as a free user.
Reserve
Cancel
vices in the same project. Learn more


Management
Disks
Networking
Access & security
External IP
Ephemeral
IP forwarding
Off
Click on Access & security


Management
Disks
Networking
Access & security
SSH Keys
Per-instance SSH keys override project-level keys. Learn more about using SSH keys.
demouser_train...
ssh-rsa AAAAB3NzaC1yc2EAAAABJQAAAQEAv6rXZJeM6N
b05eO93UFYYFkzv1++Ett5ENeGCMRDiAPXrMG6R8vZgAyD
A2VydpAl2RifEpyweL/S+UOyURtMfupg+3XzOPkf13X1nQ
bOr/zjIG/VG5Rxmxb5plJzOsgHAPNJFy0aT7Y7gqwRLbZP
w/7aV3QrGNNO0hui8y9Owq+G5SaRWD3012jDiP8N2HqBao
xvcHzLyzOgfG9Wx5A6Ec5kIW4PUQKAGvDBOkg7tOV71nbs
cHNNhoWf3/wMqkzlAvCZXMN2+NXavjZXKY7Dbmw2CwCxoL
+ Add item
Paste the contents of openSSH public.txt file here

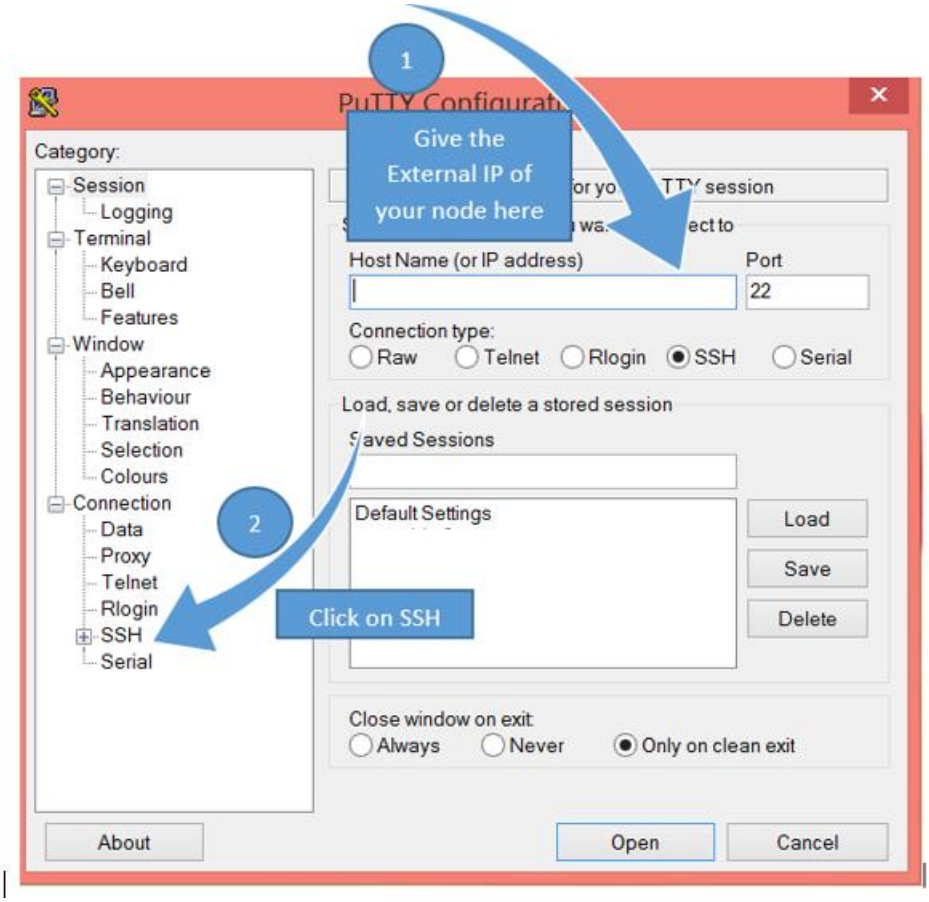


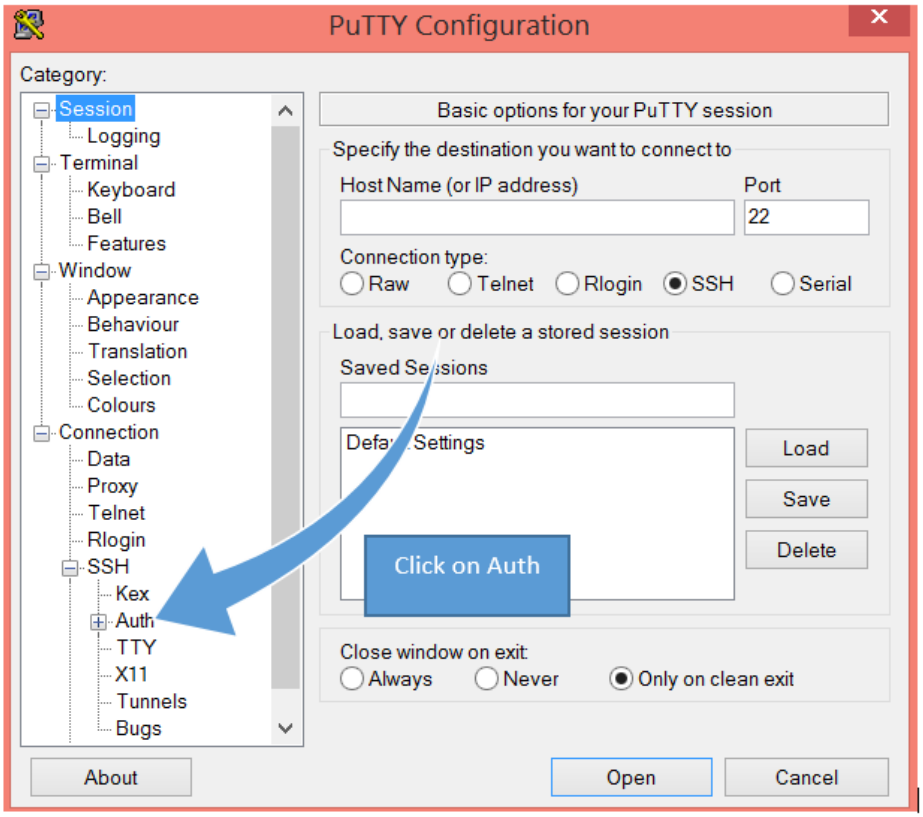
2) You have created one Virtual Machine(VM) instance on your Google cloud account.
3) This was a 2 vCPU machine. As a free user you are eligible to have only up to 8 vCPU's for an account. Ie., you can have 3 more machines with similar configuration as the one you just created. So you will be having 4 machines each with 2vCPU's.

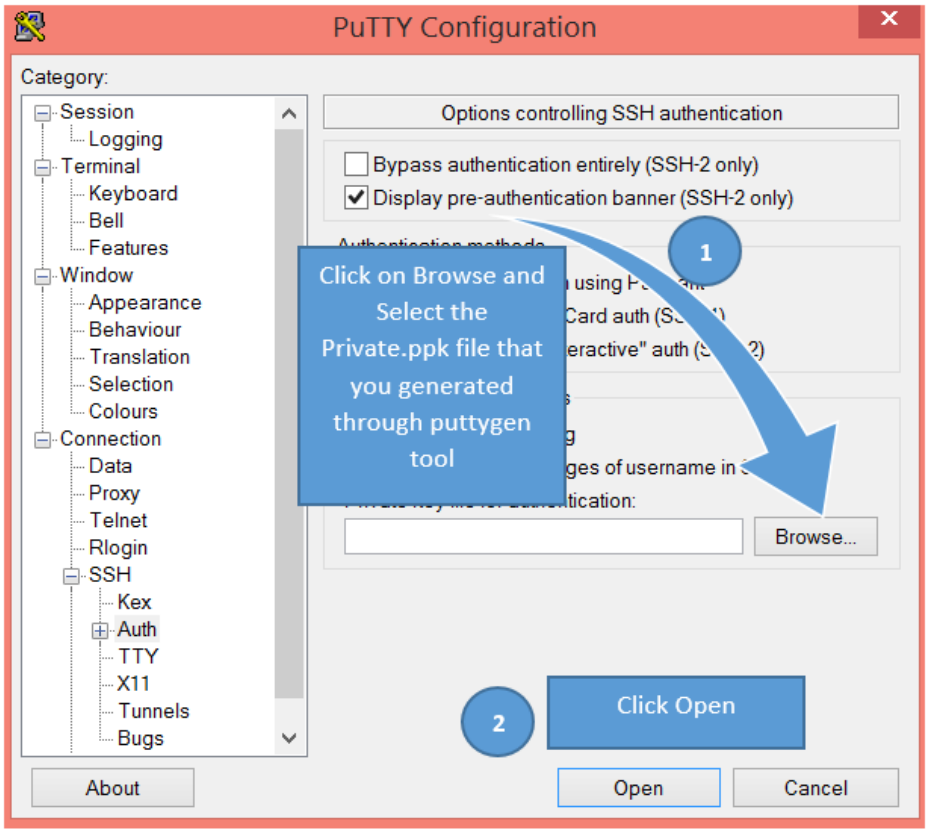
4) To create more instances, follow the Steps given in the Creating a Cluster section for each instance.
5) Remember to give the same **openSSH public** key for all the instances.
6) This ensures that all the instances can handshake with each other without a password.
7) Also as a free user you are eligible for only one Static IP, so for the other instances you create leave the IP as it is, there will be no problem.

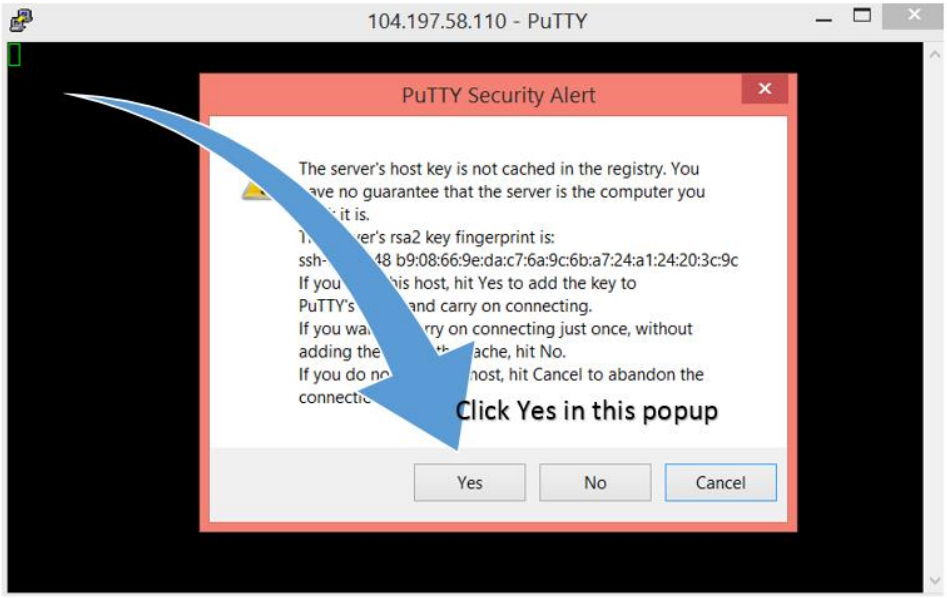
Connecting using Putty

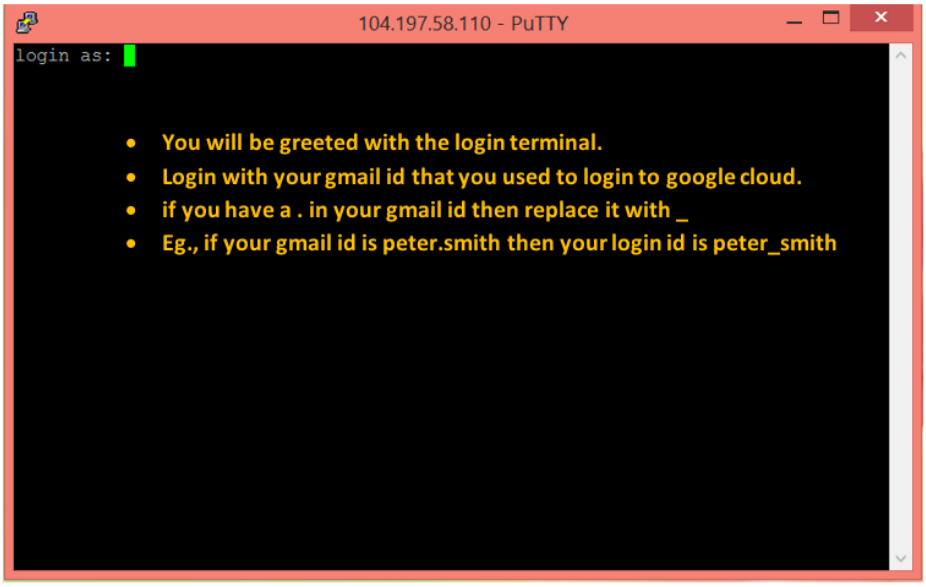
1) Download Putty.
2) Open Putty.
3) Copy the External IP of your instance (found on the dashboard page of your Google cloud).
4) This is the Master machine where we will be installing Ambari Server.
5) Copy the IP of the machine where you have given the Static IP. Let this machine which has the Static IP configured be the Master.

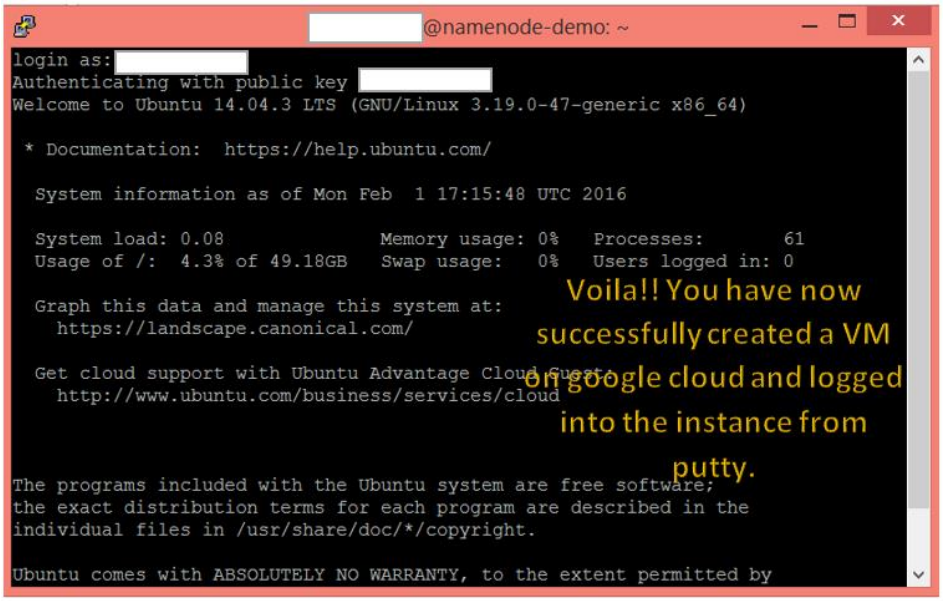



PuTTY Configuration
Category:
Session
Logging
Terminal
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
Telnet
Rlogin
SSH
Kex
Auth
TTY
X11
Tunnels
Bugs
Basic options for your PuTTY session
Specify the destination you want to connect to
Host Name (or IP address)
Port
22
Connection type:
Raw
Telnet
Rlogin
SSH
Serial
Load, save or delete a stored session
Saved Sessions
Load
Save
Delete
Click on Auth
Close window on exit:
Always
Never
Only on clean exit
About
Open
Cancel


PuTTY Configuration
Category:
Session
Logging
Terminal
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
Telnet
Rlogin
SSH
Kex
Auth
TTY
X11
Tunnels
Bugs
Options controlling SSH authentication
Bypass authentication entirely (SSH-2 only)
Display pre-authentication banner (SSH-2 only)
1
Click on Browse and Select the Private.ppk file that you generated through puttygen tool
Browse...
2
Click Open
About
Open
Cancel

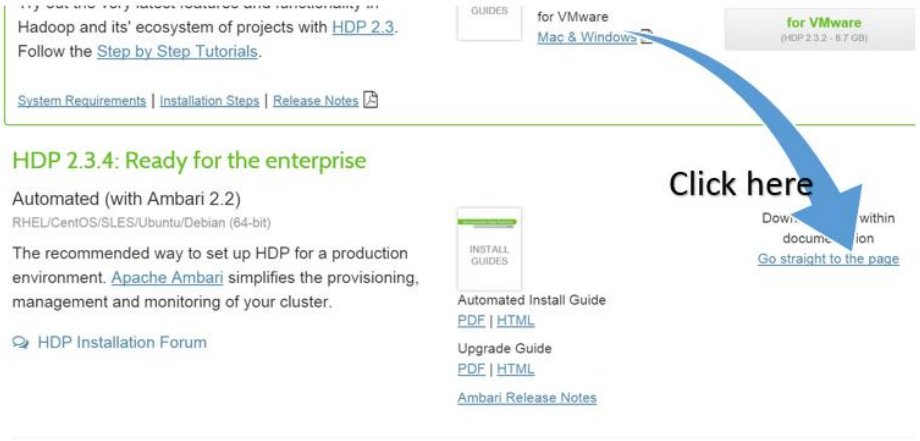

104.197.58.110 - PuTTY
PuTTY Security Alert
The server's host key is not cached in the registry. You
ave no guarantee that the server is the computer you
it is.
er's rsa2 key fingerprint is:
48 b9:08:66:9e:da:c7:6a:9c:6b:a7:24:a1:24:20:3c:9c
is host, hit Yes to add the key to
and carry on connecting.
rry on connecting just once, without
ache, hit No.
host, hit Cancel to abandon the
Click Yes in this popup
Yes
No
Cancel

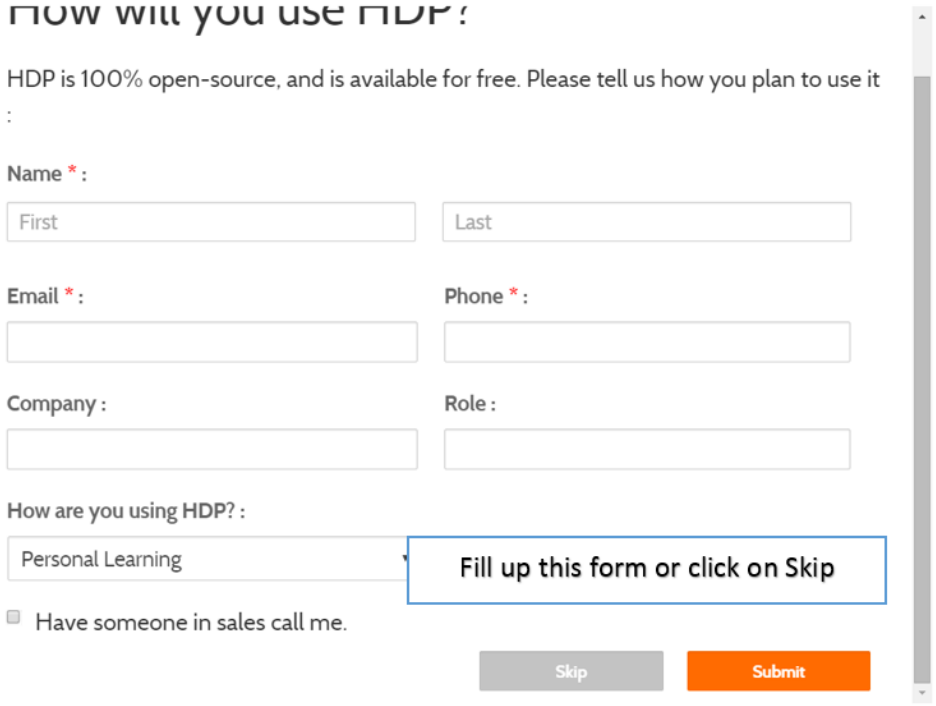

104.197.58.110 - PuTTY
login as:
You will be greeted with the login terminal.
Login with your gmail id that you used to login to google cloud.
if you have a . in your gmail id then replace it with _
Eg., if your gmail id is peter.smith then your login id is peter_smith

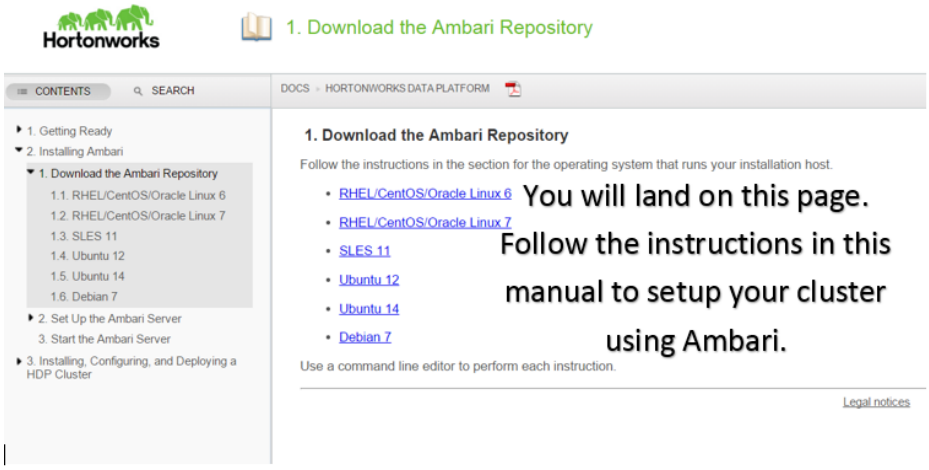

@namenode-demo: ~
login as:
Authenticating with public key
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.19.0-47-generic x86_64)
* Documentation: https://help.ubuntu.com/
System information as of Mon Feb 1 17:15:48 UTC 2016
System load: 0.08
Memory usage: 0%
Processes: 61
Usage of /: 4.3% of 49.18GB
Swap usage: 0%
Users logged in: 0
Graph this data and manage this system at:
https://landscape.canonical.com/
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud
Voila!! You have now successfully created a VM on google cloud and logged into the instance from putty.
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by

Ambari Setup

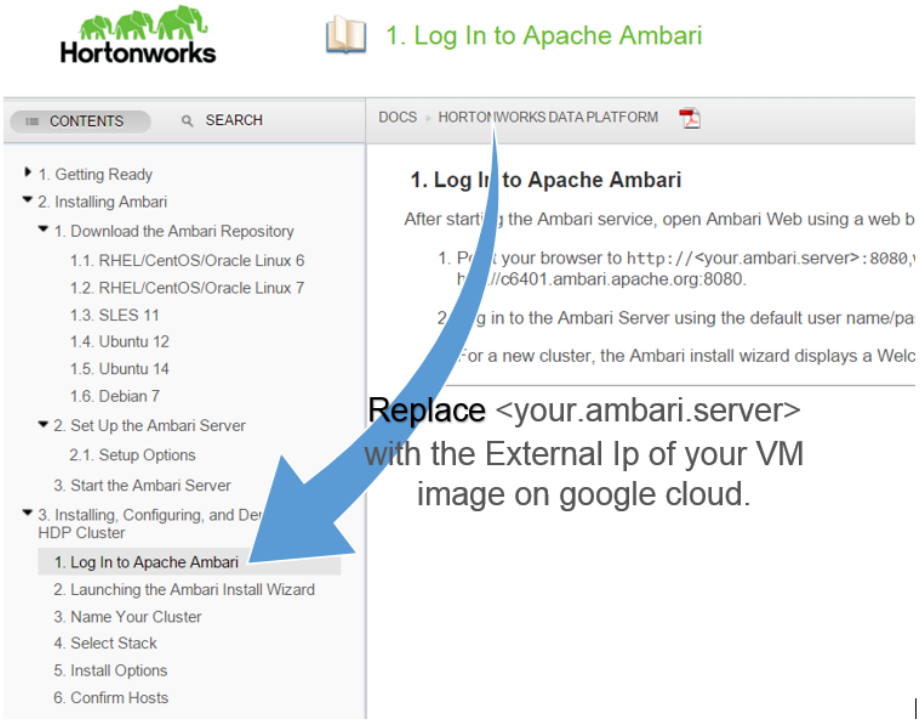
1) Go to http://hortonworks.com/hdp/downloads/
2) This is the Horton Works download page which contains download instructions for different modes of installing HDP.
3) We are going to setup the HDP cluster on our Google cloud instance using Ambari.
4) Follow the steps below.



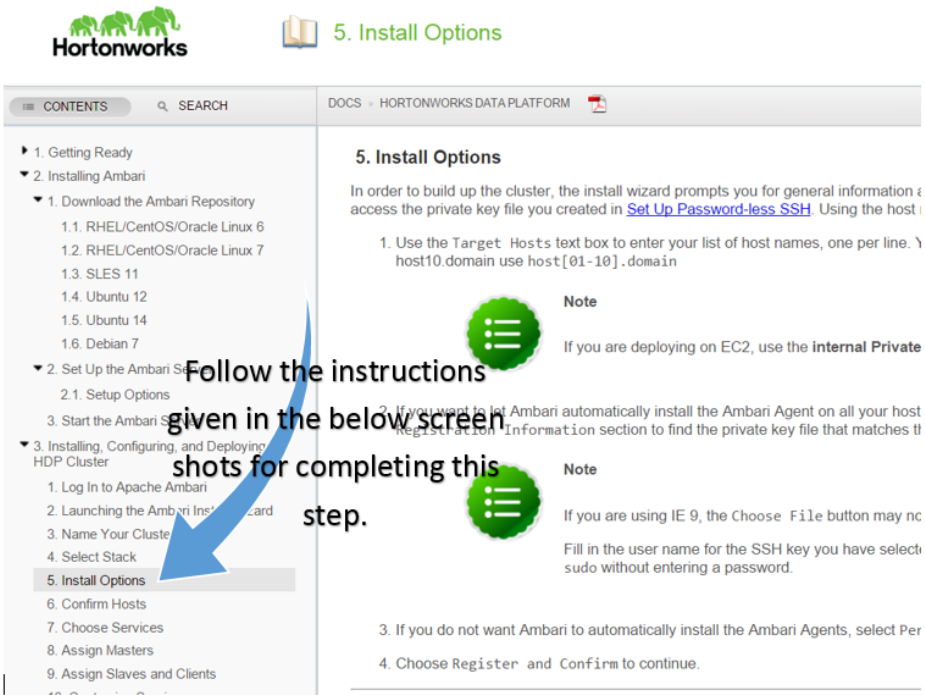




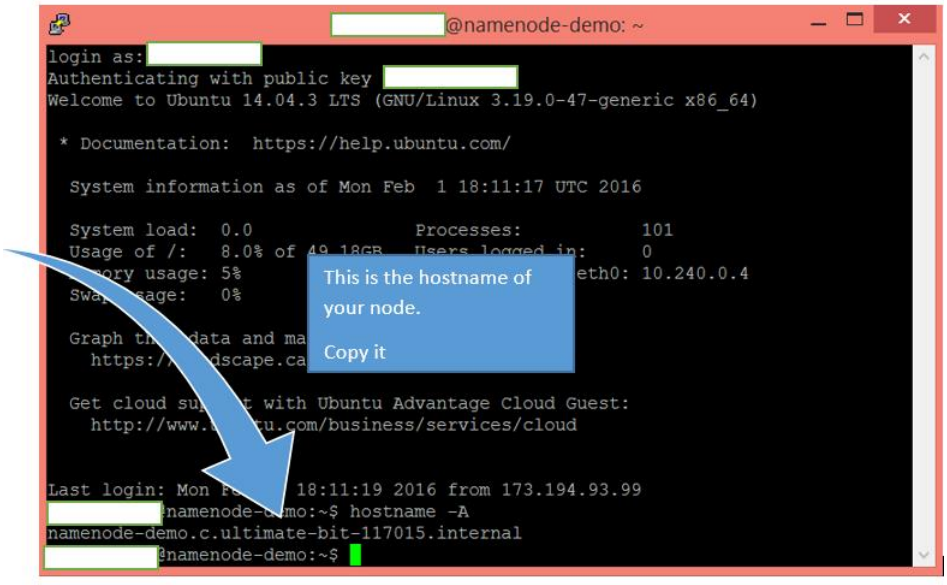
5) use **'sudo'** before running any command given in the instruction manual.
6) Follow the steps given in the manual until you reach the below step.

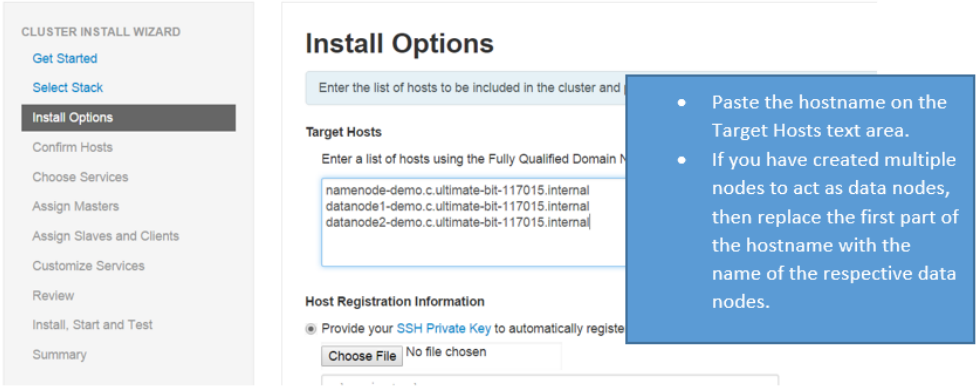


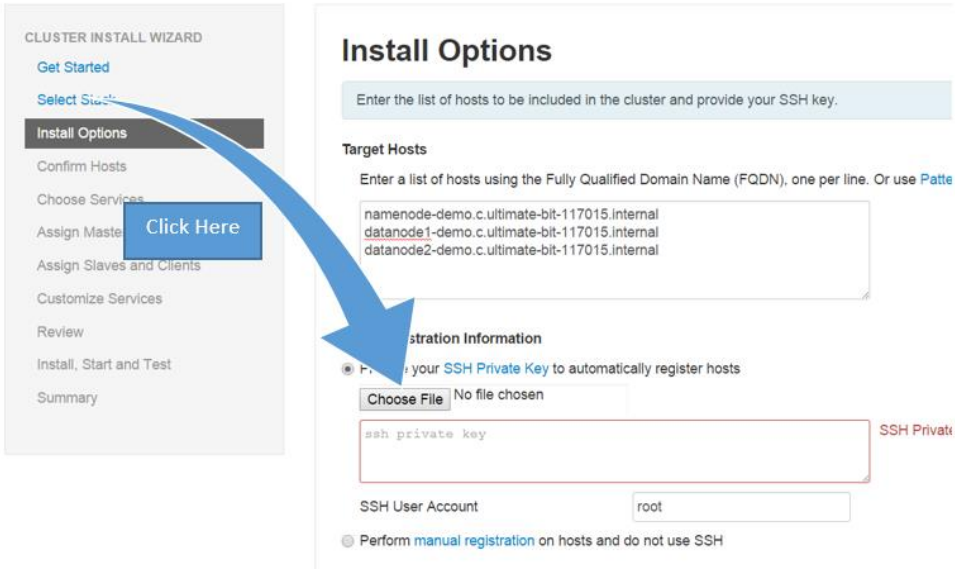
7) Follow the steps given in the manual until you reach the below step.

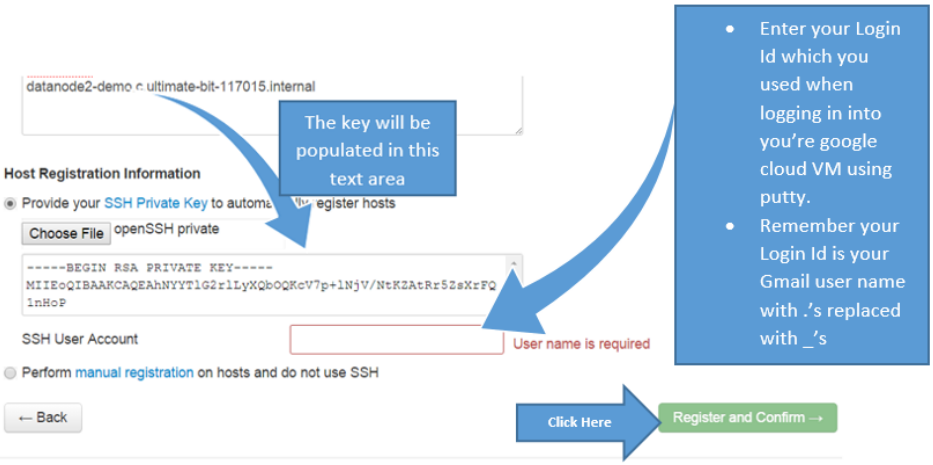


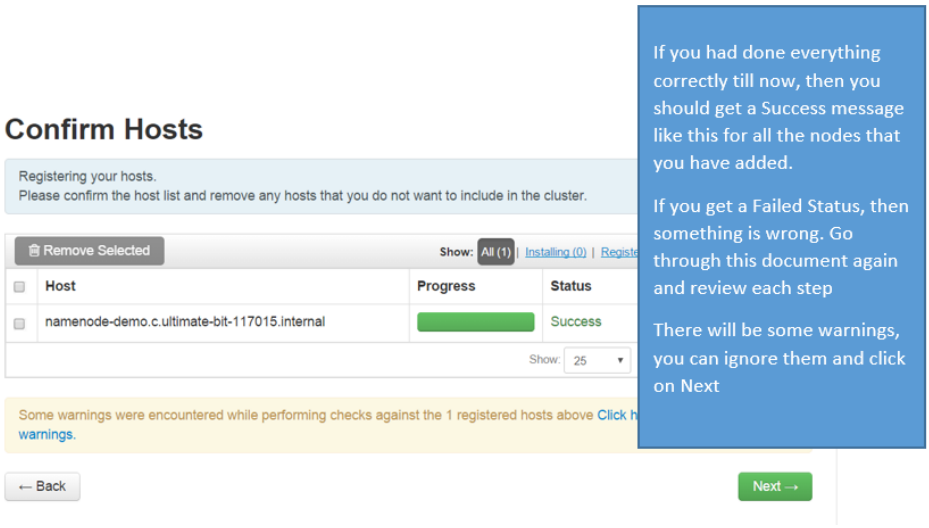
8) Open your putty window and enter the command **hostname -A**






CLUSTER INSTALL WIZARD
Get Started
Select Slaves
Install Options
Confirm Hosts
Choose Services
Assign Masters
Assign Slaves and Clients
Customize Services
Review
Install, Start and Test
Summary
Click Here
Install Options
Enter the list of hosts to be included in the cluster and provide your SSH key.
Target Hosts
Enter a list of hosts using the Fully Qualified Domain Name (FQDN), one per line. Or use Patte
namenode-demo.c.ultimate-bit-117015.internal
datanode1-demo.c.ultimate-bit-117015.internal
datanode2-demo.c.ultimate-bit-117015.internal
stration Information
your SSH Private Key to automatically register hosts
Choose File
No file chosen
ssh private key
SSH Privat
SSH User Account
root
Perform manual registration on hosts and do not use SSH


datanode2-demo.c.ultimate-bit-117015.internal
The key will be populated in this text area
Host Registration Information
Provide your SSH Private Key to automa
egister hosts
Choose File
openSSH private
-----BEGIN RSA PRIVATE KEY-----
MIIEoQIBAAKCAQEAhNYYTlG2rlLyXQbOQKcV7p+lNjV/NtKZAtRr5ZsXrFQ
lnHoP
SSH User Account
User name is required
Perform manual registration on hosts and do not use SSH
← Back
Click Here
Register and Confirm →
Enter your Login Id which you used when logging in into you're google cloud VM using putty.
Remember your Login Id is your Gmail user name with .'s replaced with _'s



9) Continue the steps given on the Horton Works Ambari Installation manual to complete the configuration of your Hadoop cluster.

"All the best"