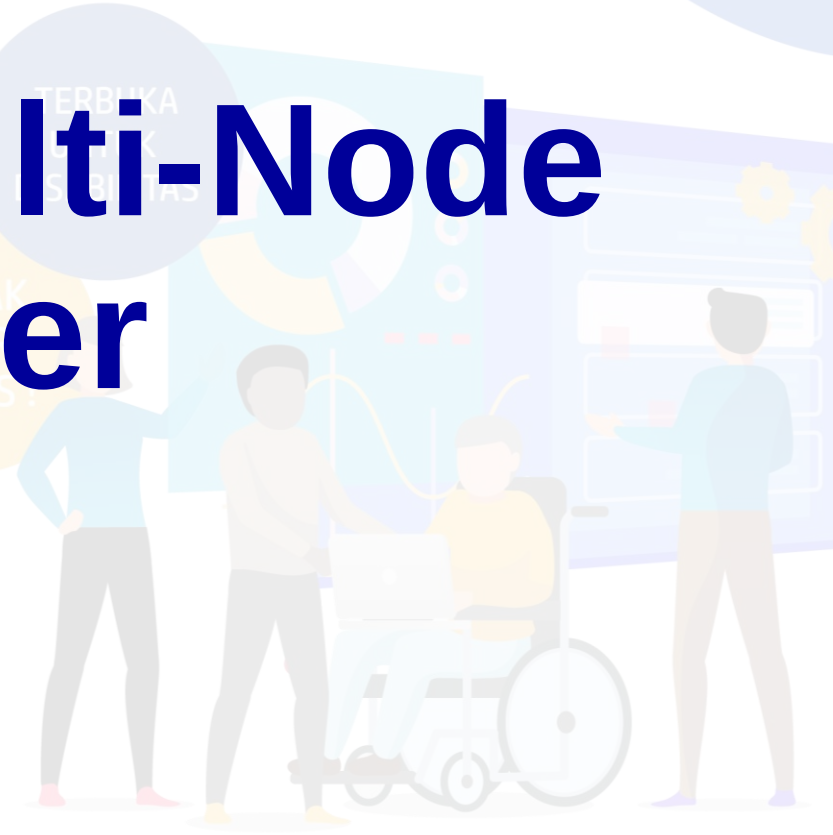
DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics

digitalent.kominfo.go.id

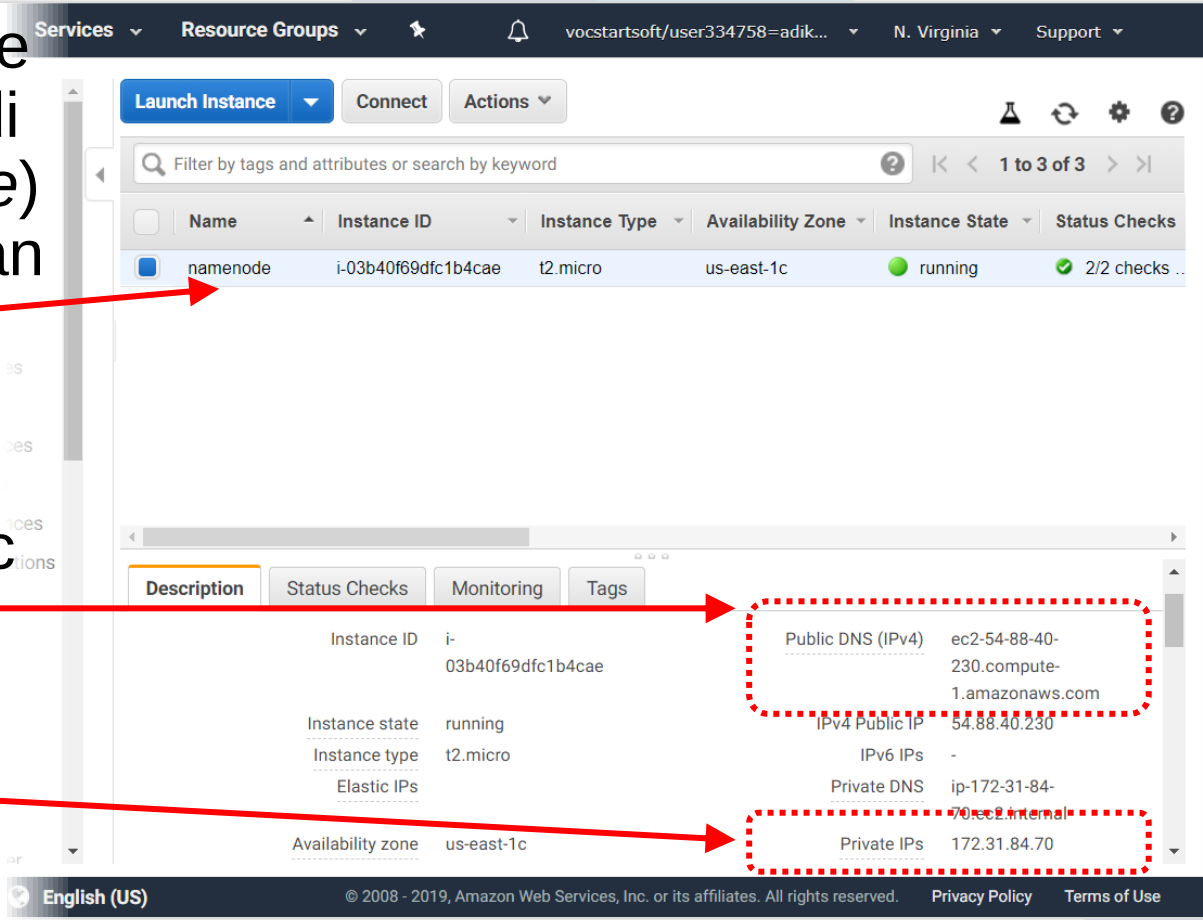
Hadoop Multi-Node Cluster



Membuat Instance untuk Master & Slave

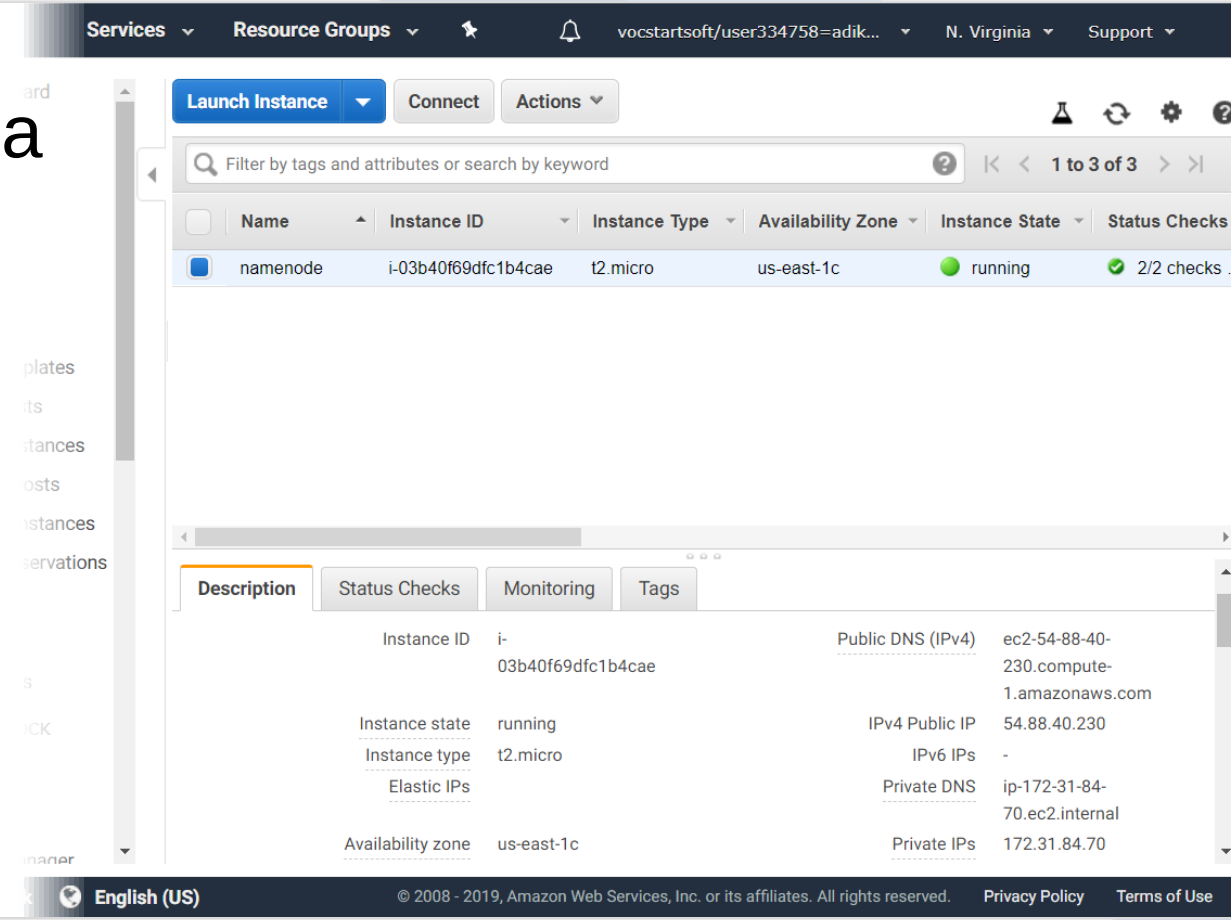
Membuat Master (Namenode)

- Beri nama instance yang akan menjadi master (*namenode*) untuk memudahkan pengelolaan
  - Saat hover, klik tombol pensil
- Catat (*copy*) public DNS:
- ec-2-x-x-x.compute1…
- Catat Private IPs



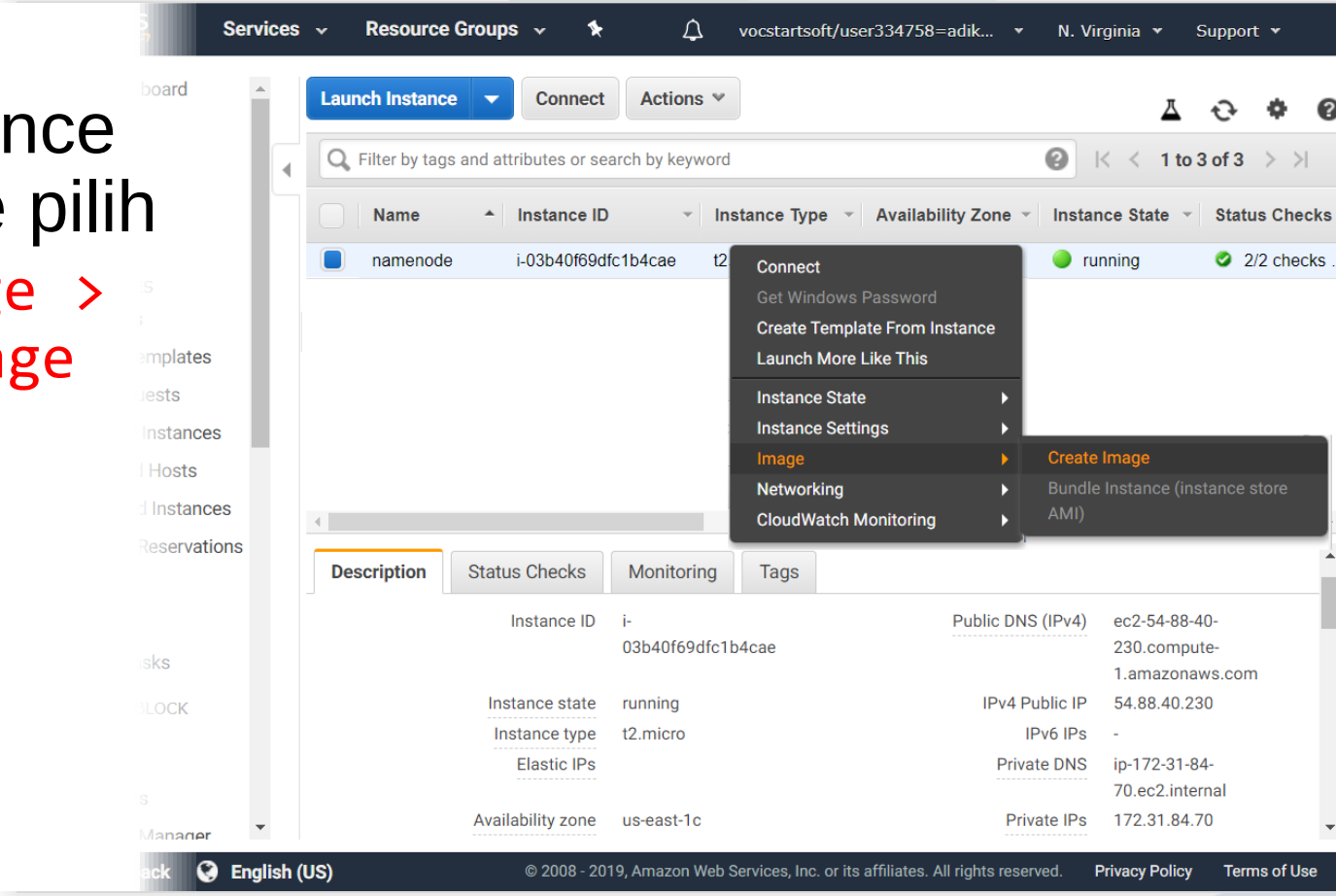
Membuat Slave (Datanode) (1)

- Dari instance yang sudah ada yang menjadi master/ namenode maka buat Image baru untuk menjadi datanode

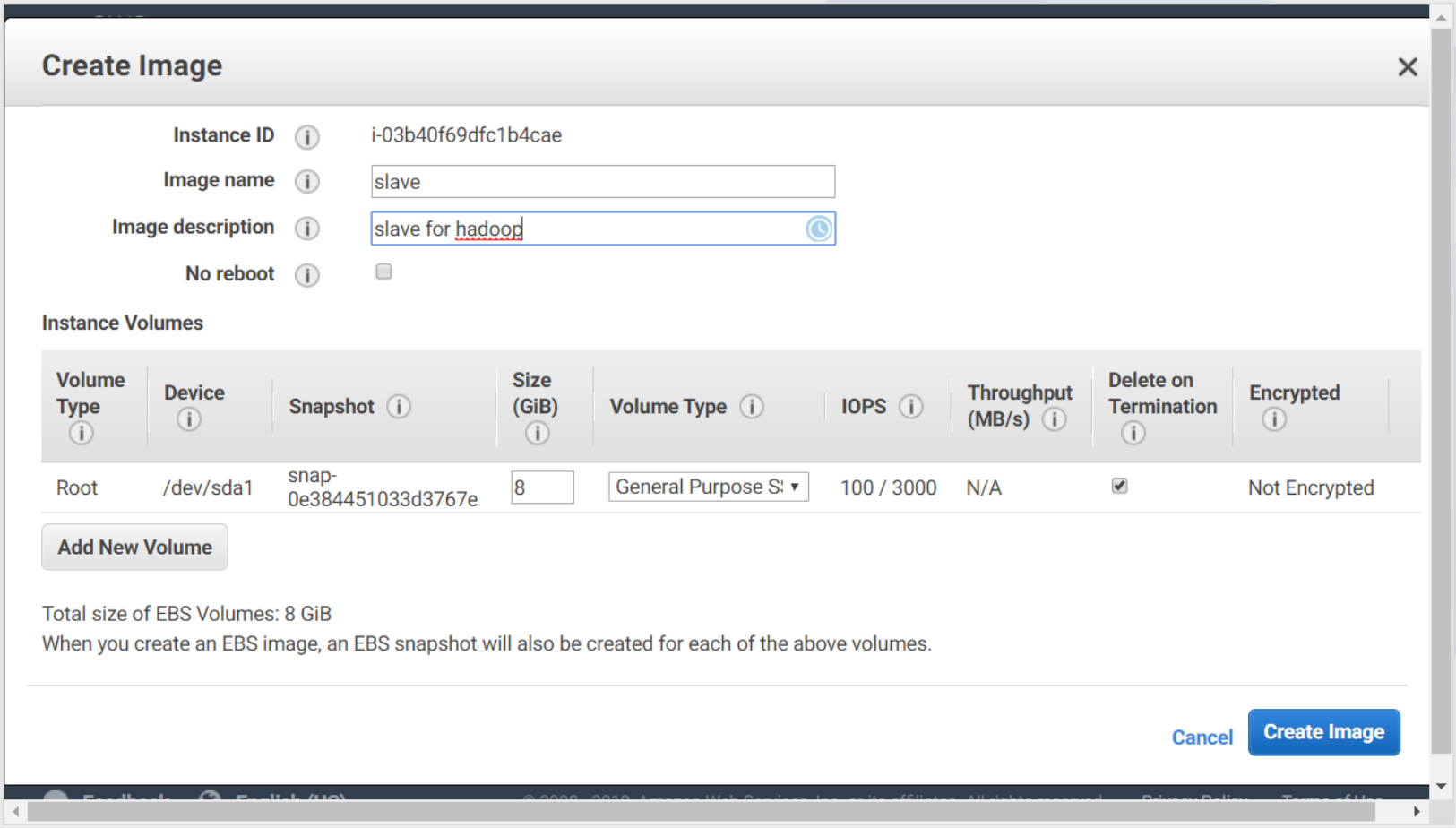


Membuat Slave (Datanode) (2)

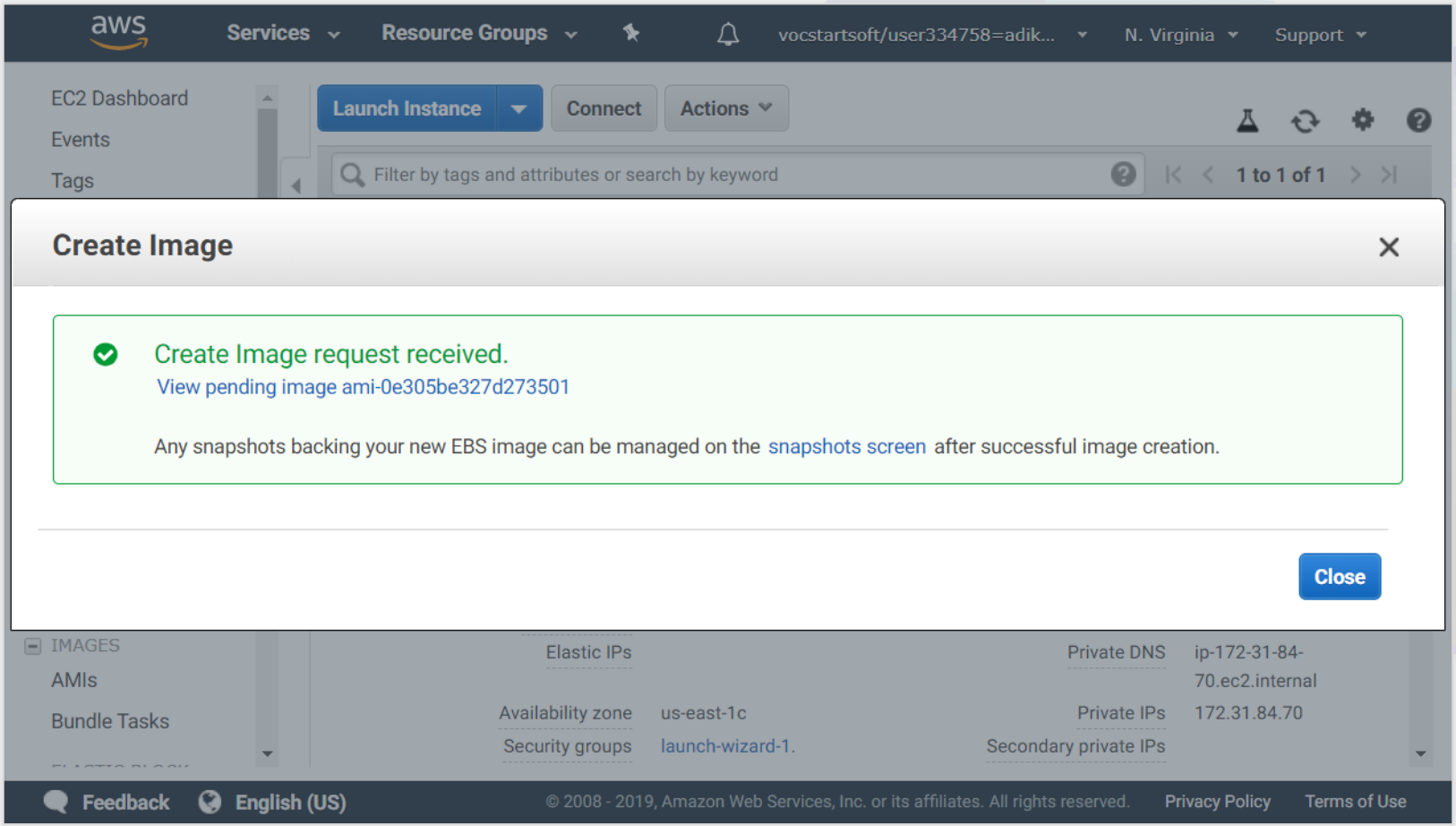
- Klik kanan pada instance namenode pilih menu `Image > Create Image`



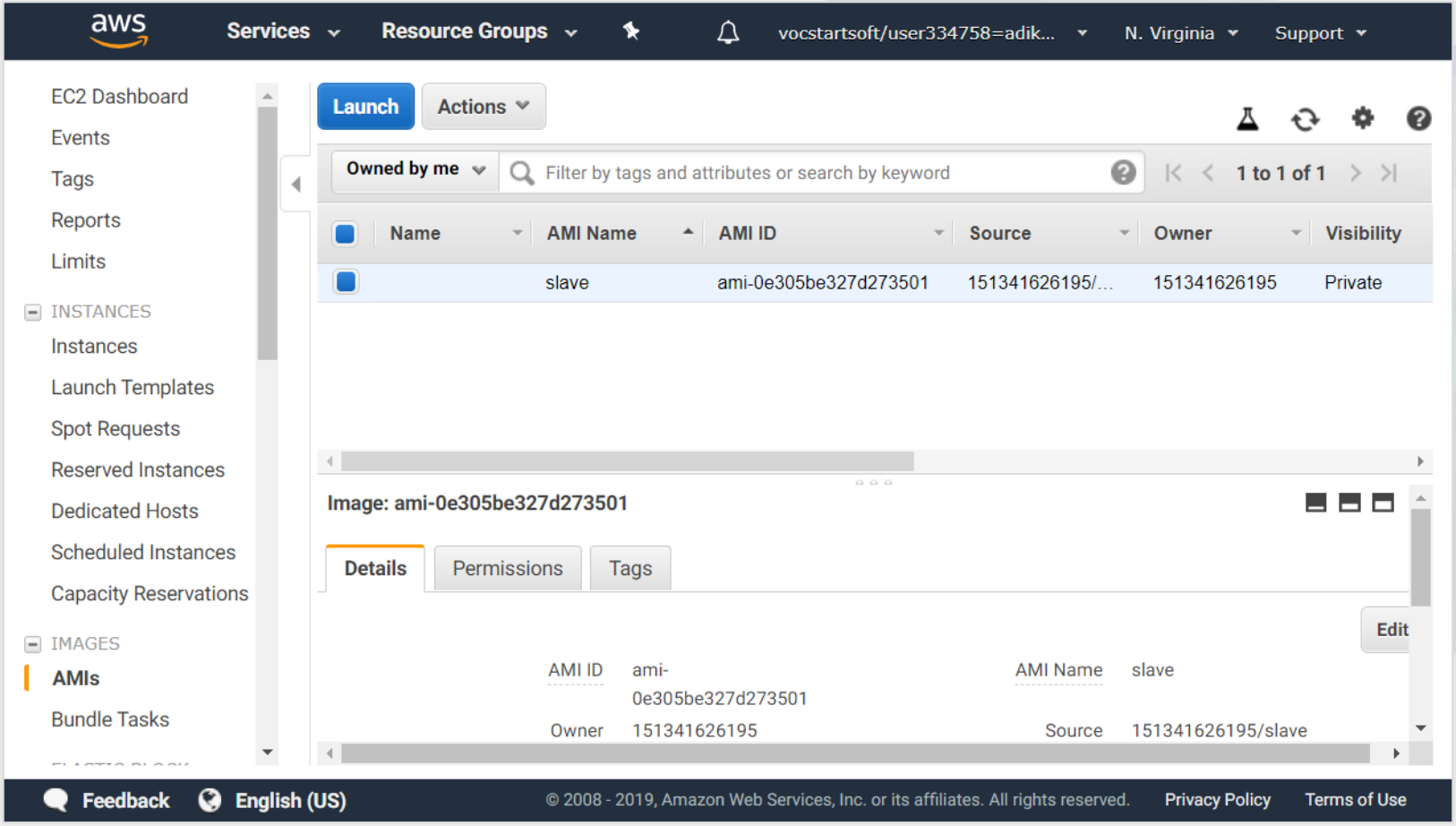
Membuat Slave (Datanode) (3)



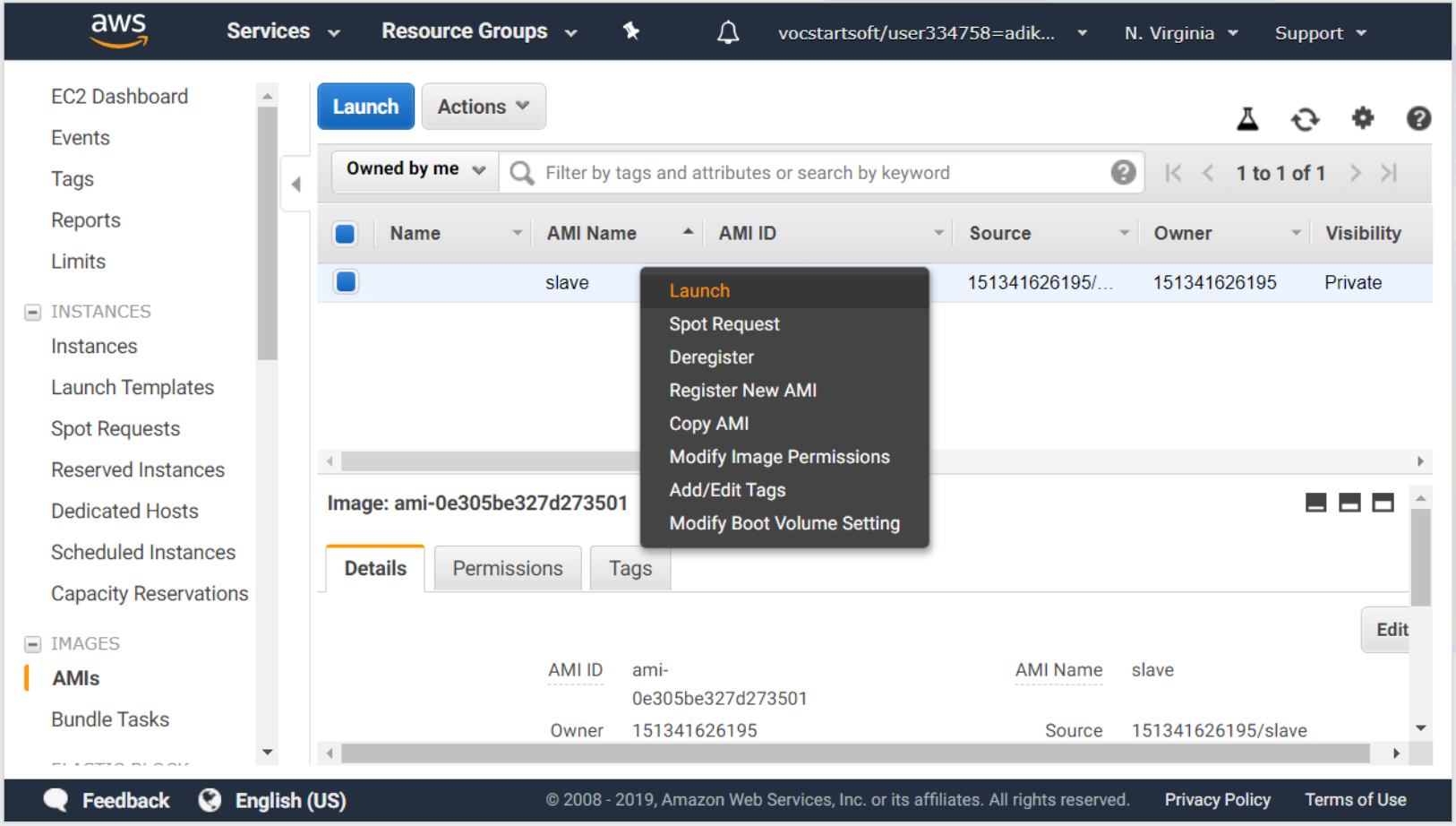
Membuat Slave (Datanode) (4)



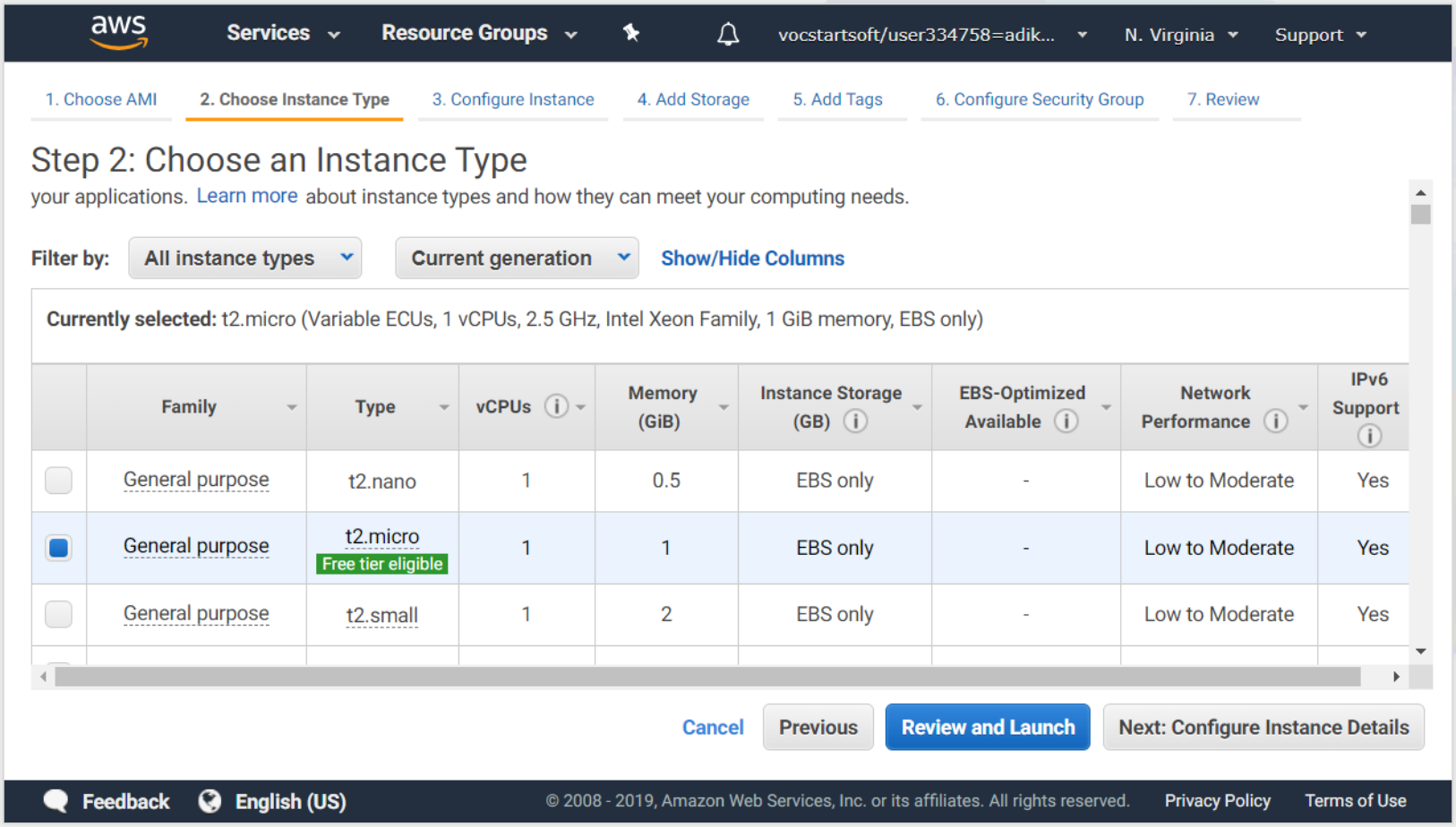
Membuat Slave (Datanode) (5)



Membuat Slave (Datanode) (6)

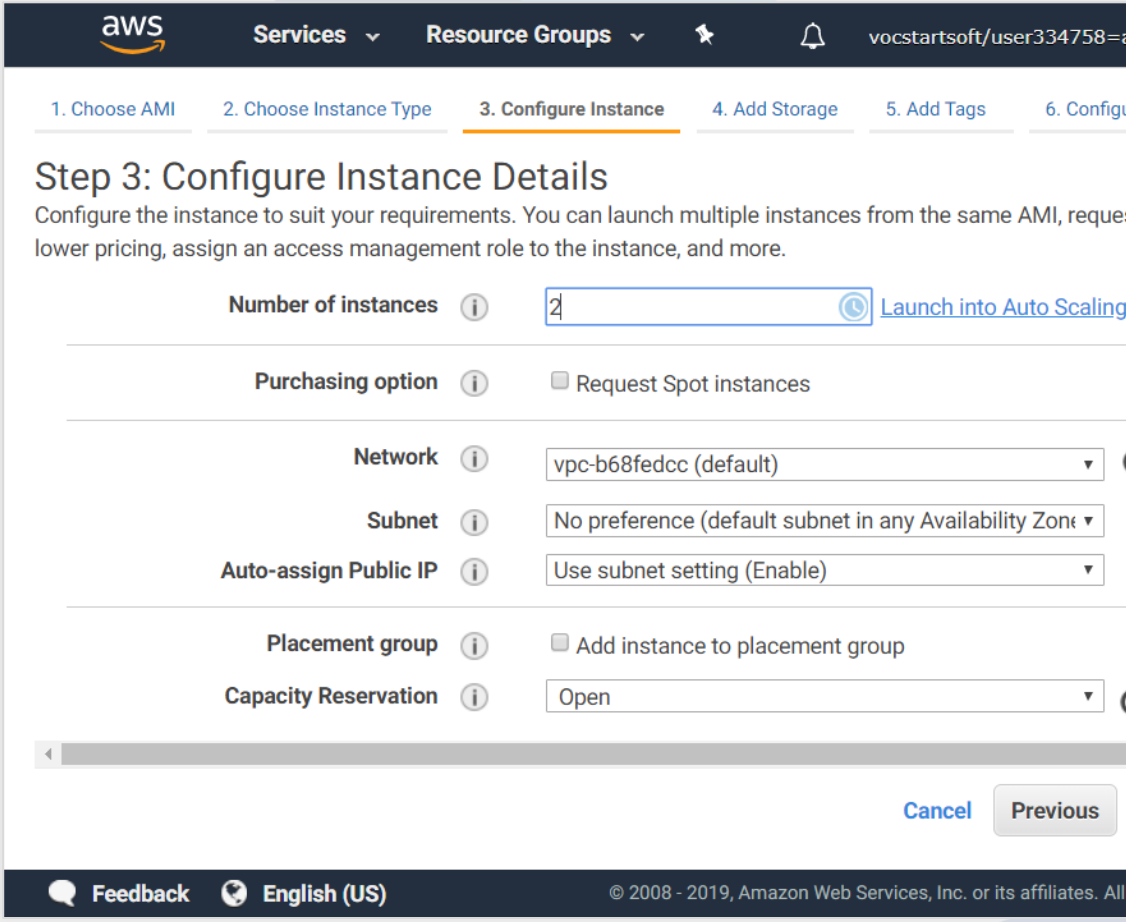


Membuat Slave (Datanode) (7)

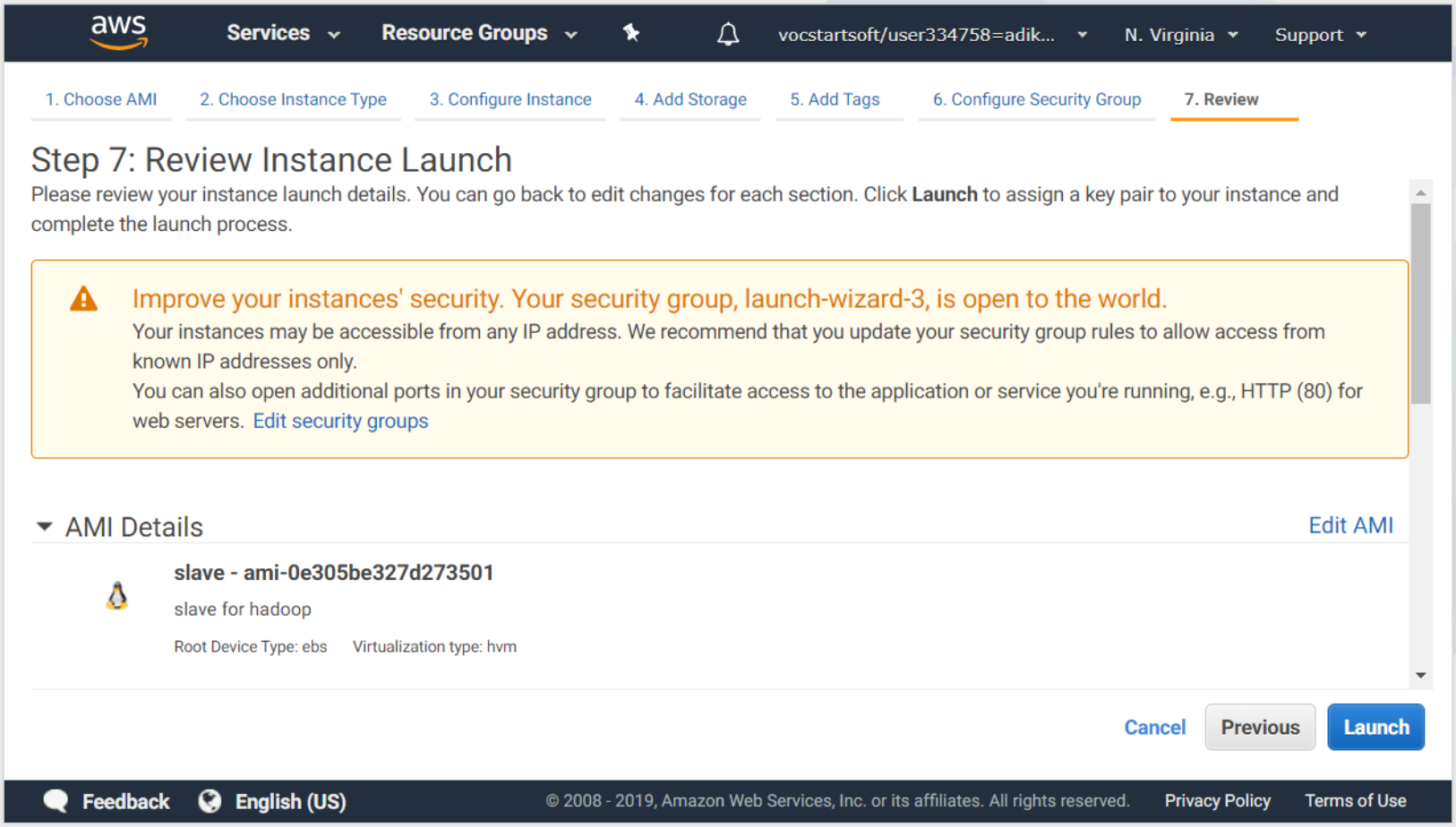


Membuat Slave (Datanode) (8)

- Isi **Number of instance** sebanyak datanode yang ingin dibuat
- Misalnya dalam contoh buat 2 datanode

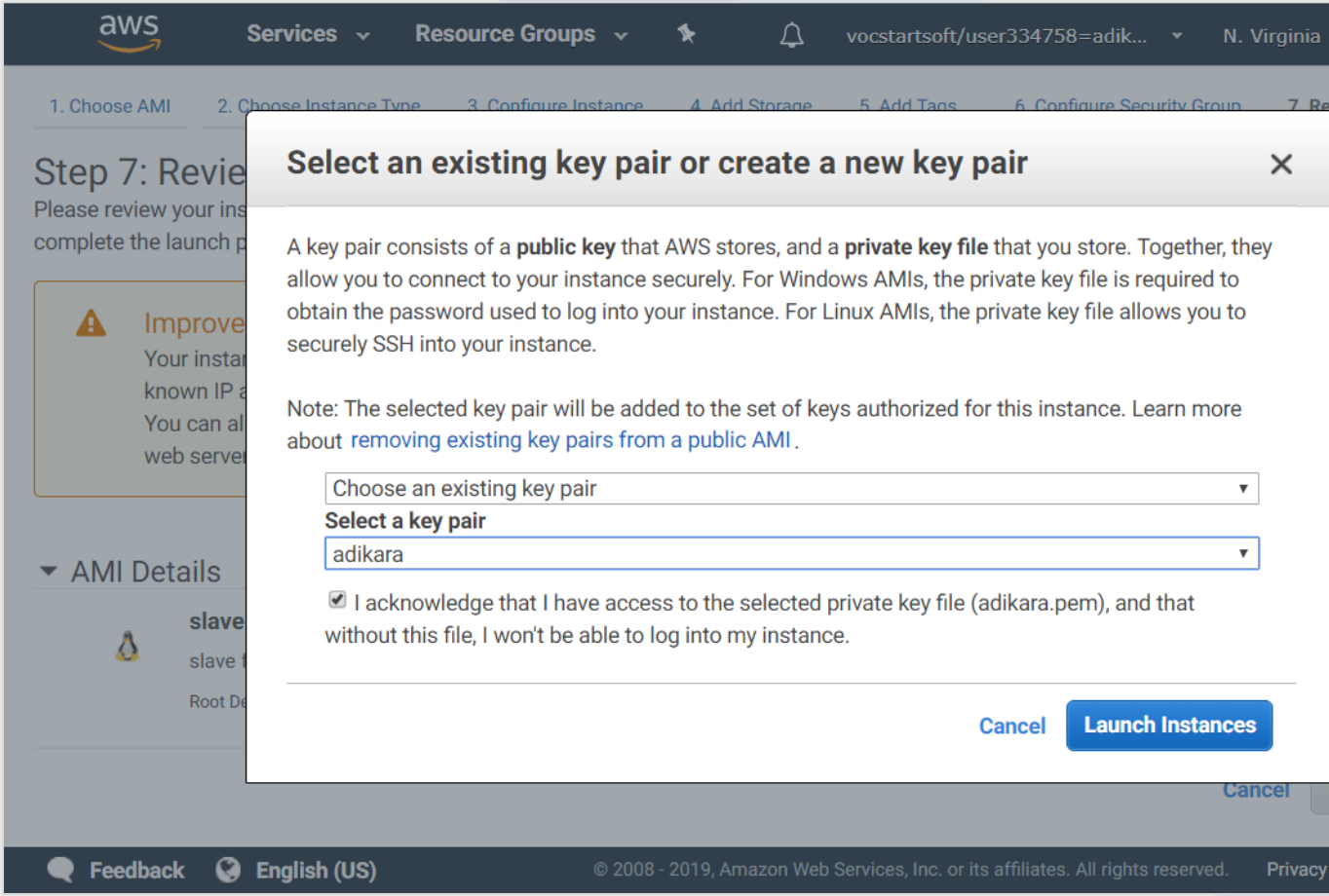


Membuat Slave (Datanode) (9)

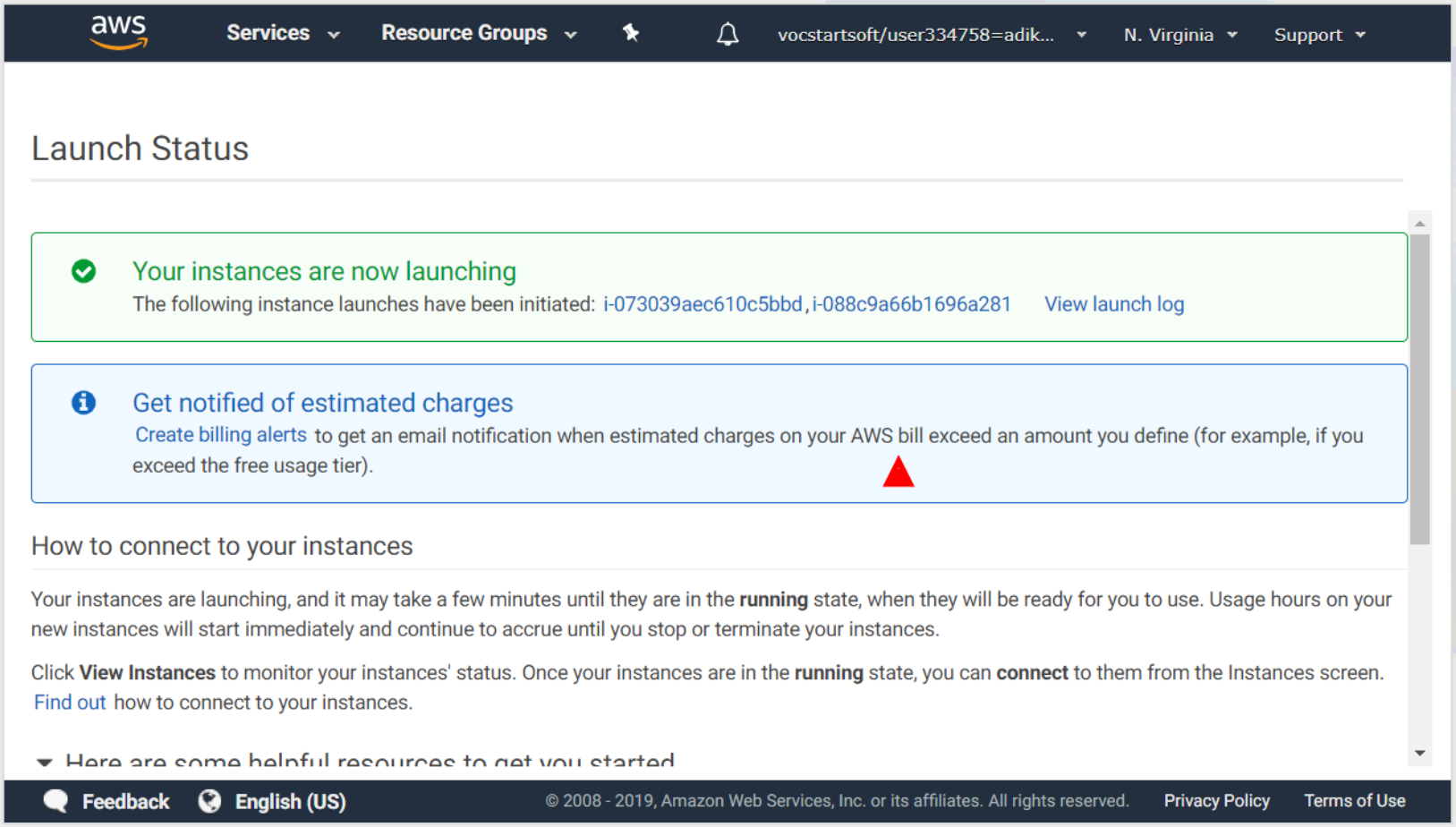


Membuat Slave (Datanode) (10)

- Gunakan key pair yang sama dengan yang dibuat untuk namenode

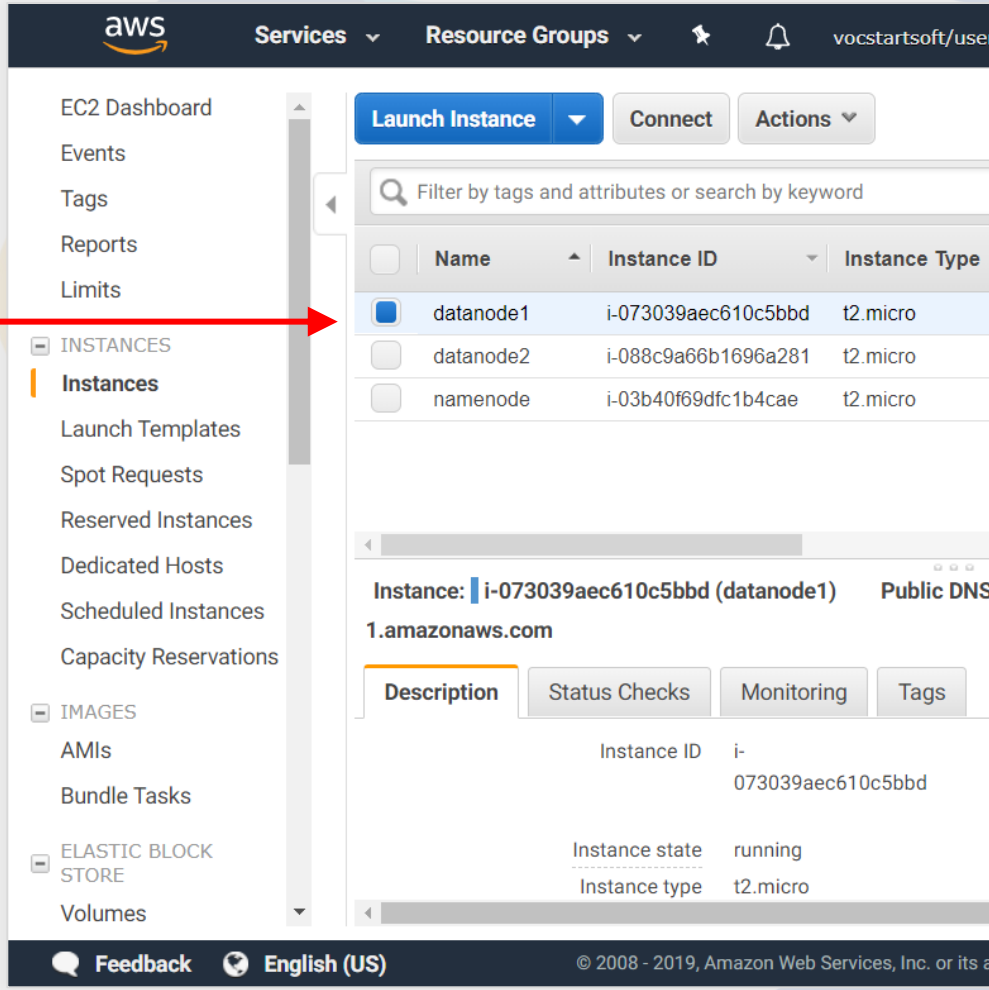


Membuat Slave (Datanode) (12)



Membuat Slave (Datanode) (13)

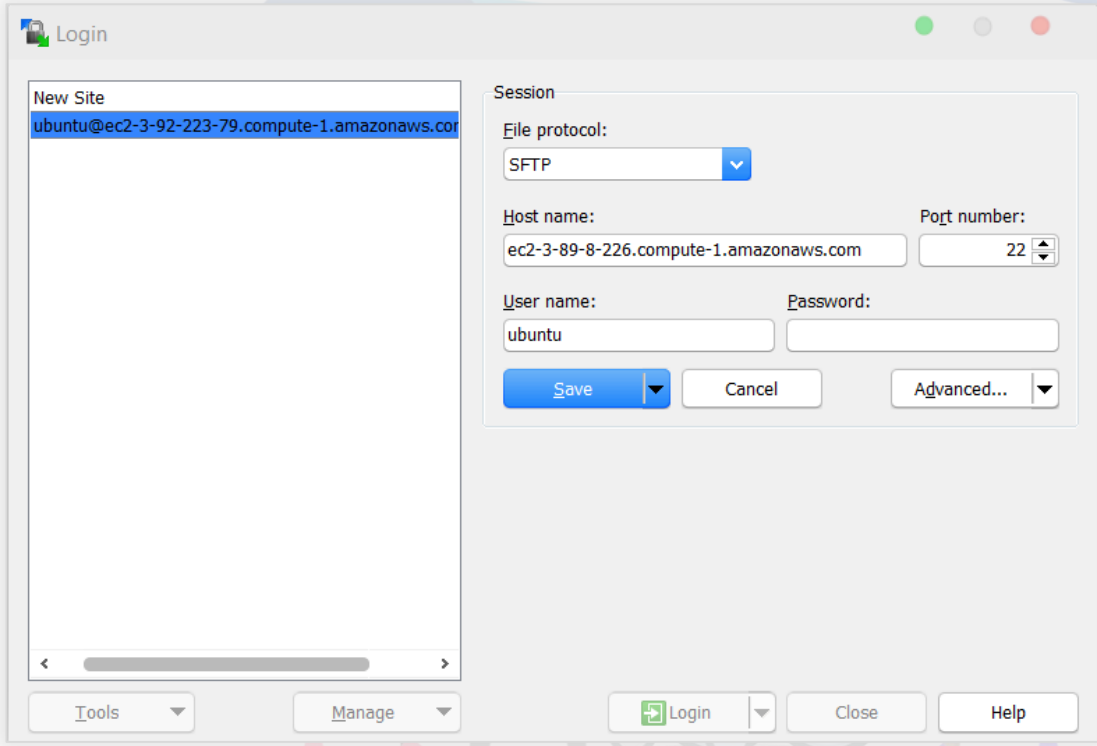
- Ubah nama masing-masing instance dari image yang baru dibuat menjadi datanode1, datanode2, dst.



Buat Koneksi SSH

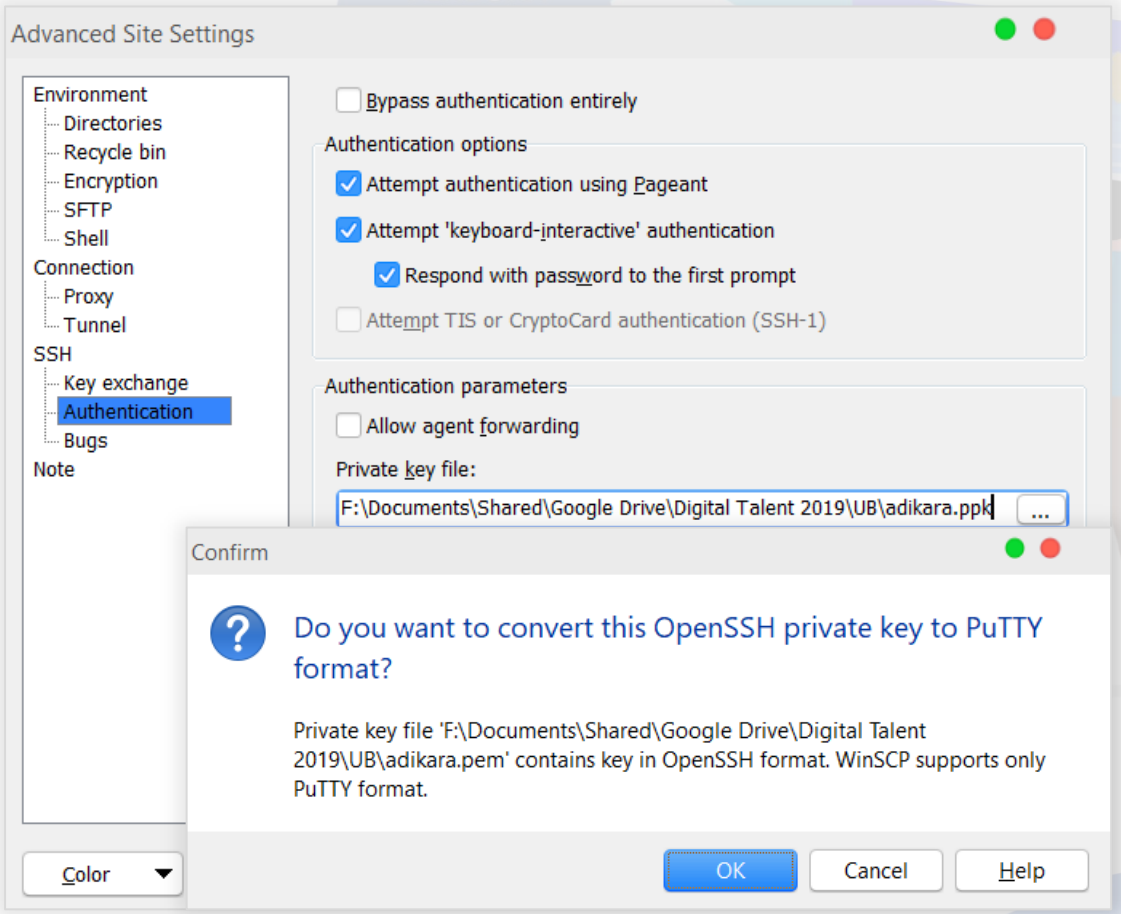
Buat Koneksi dengan WinSCP (1)

- Buka WinSCP
- Isikan hostname sesuai Public DNS di EC2
- Isikan User name = "ubuntu"
- Klik tombol Advanced…



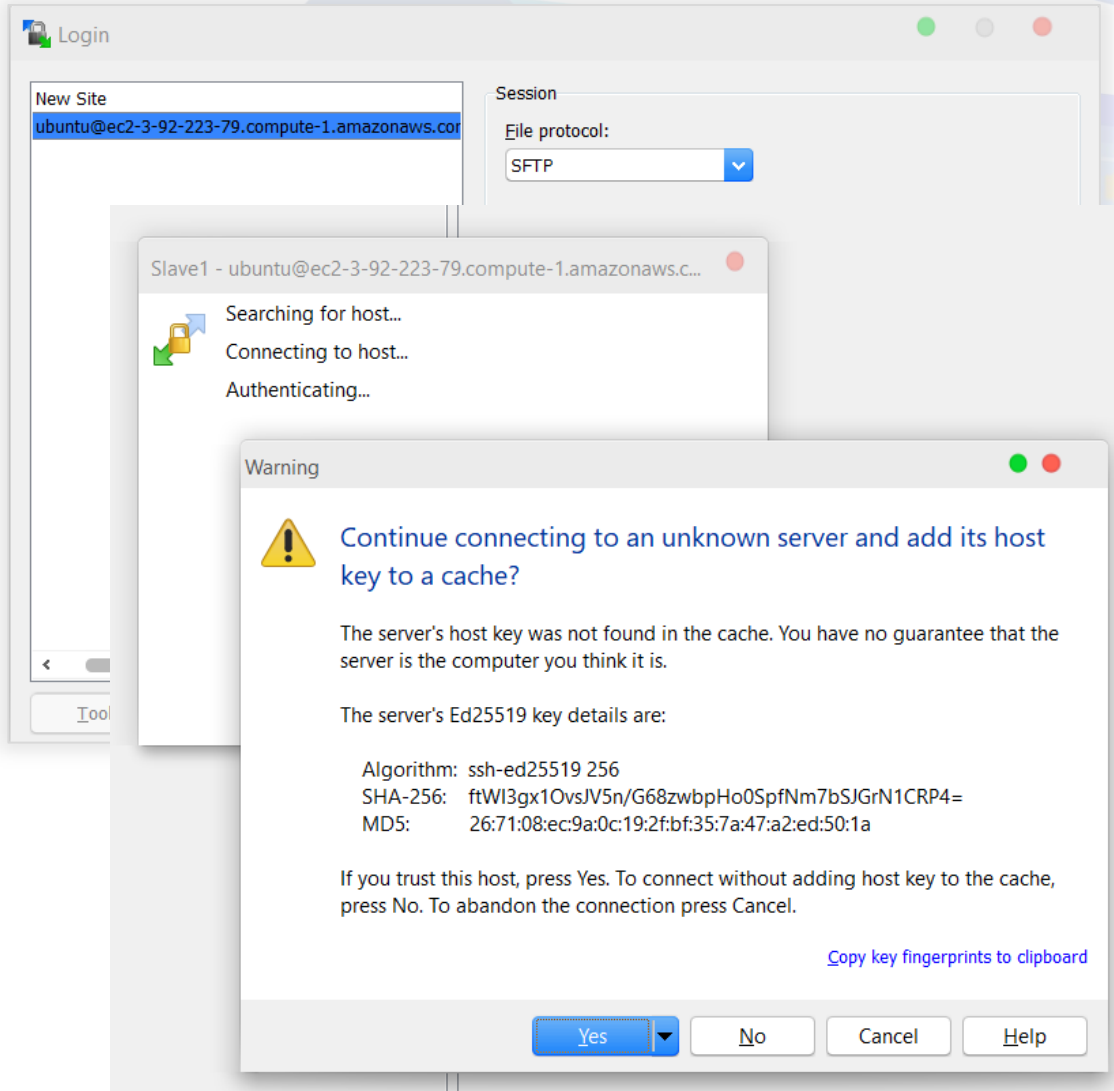
Buat Koneksi dengan WinSCP (2)

- Pilih SSH > Authentication
- Pilih fail .pem yang sudah diunduh
- Ketika diminta mengubah dari fail OpenSSH (.pem) ke PuTTY (.ppk) pilih OK
- Tutup jendela Advanced dengan pilih OK



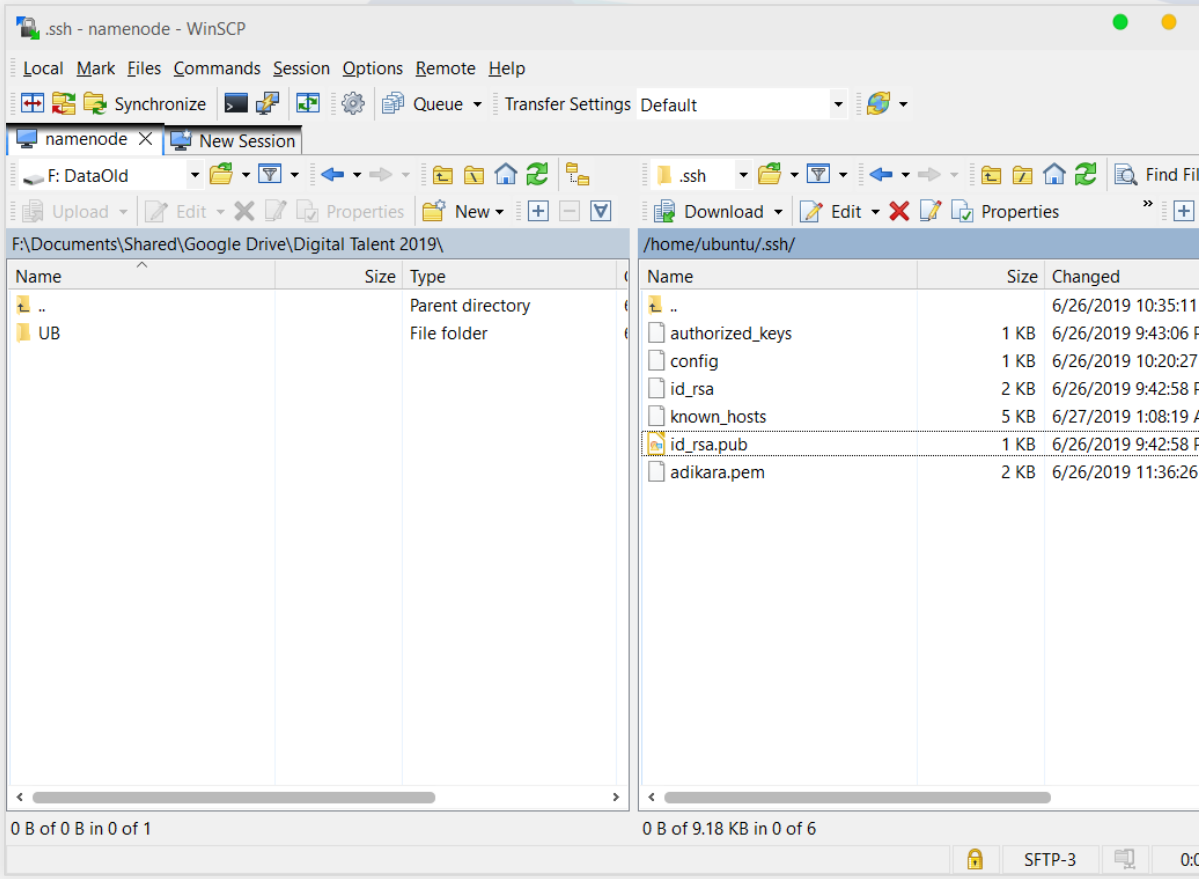
Buat Koneksi dengan WinSCP (3)

- Klik Login
- Apabila muncul jendela Warning pilih Yes



Buat Koneksi dengan WinSCP (4)

- Di panel sebelah kanan pilih folder `~/.ssh`
- Misalnya bila nama user=`ubuntu`, maka masuk ke folder `/home/ubuntu/.ssh`

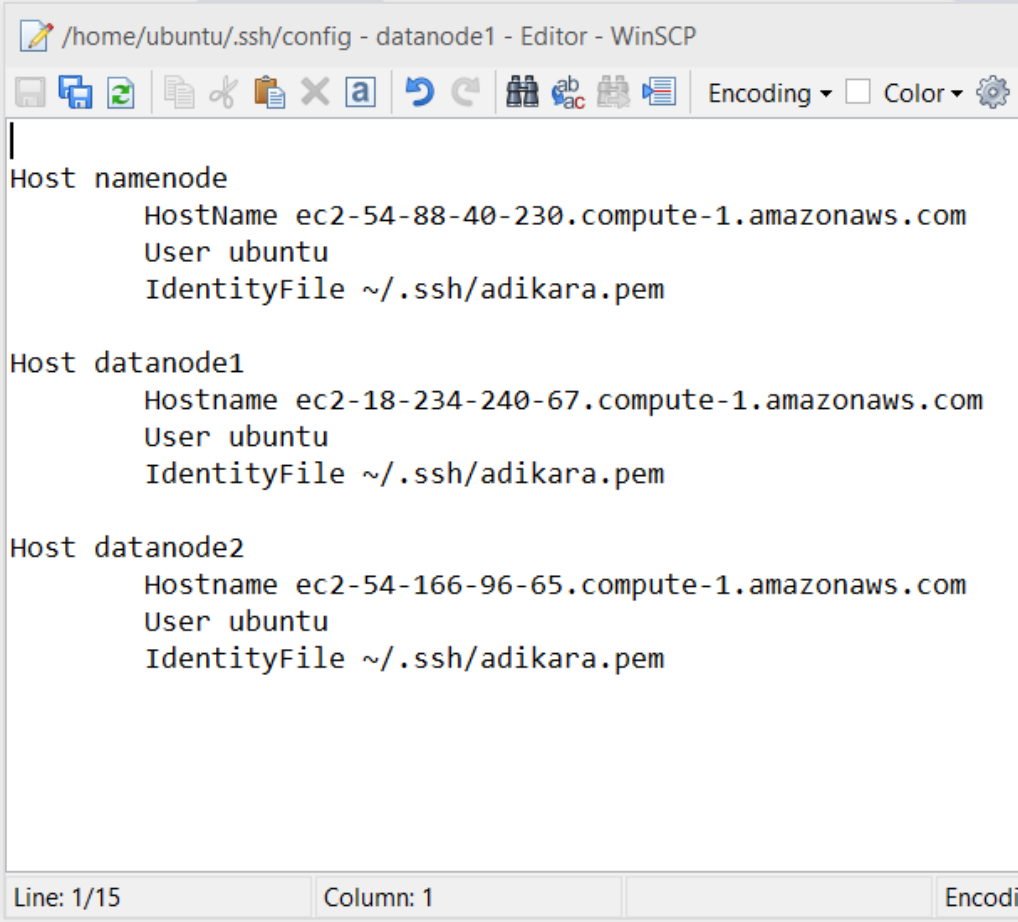


Setting Password-less SSH (1)

- Melalui WinSCP (atau PuTTY) kita menggunakan koneksi dengan password (melalui fail .pem)
- Supaya Hadoop bisa terhubung secara otomatis dari namenode ke semua datanode (dan sebaliknya) maka diperlukan koneksi SSH tanpa password (*password-less)*

Setting SSH di Semua Node

- Buat atau ubah fail `config` di `~/.ssh/config`
  - Di panel kanan klik kanan, `New > File`
- Buat entry untuk semua node (namenode & datanode)
  - Beri nama `Host` sesuai fungsinya
  - Nama `Hostname` masing-masing host adalah public DNS
  - User=ubuntu
  - Tentukan lokasi IdentityFile (.pem)

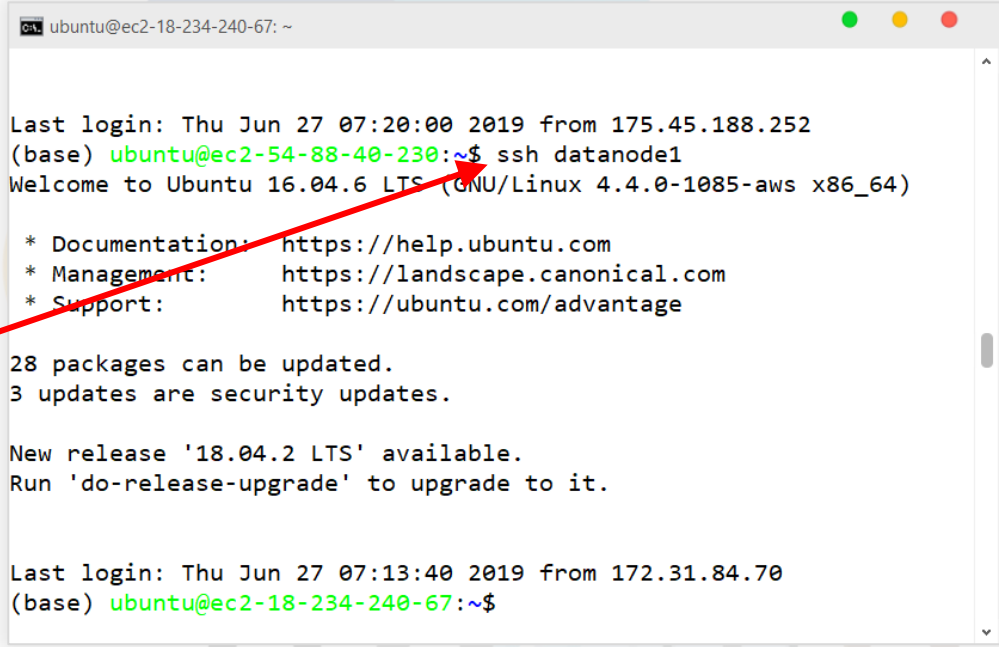


Kopikan Public Key ke Semua Node

- Login dengan PuTTY ke namenode atau lewat SSH
  - `ssh -i "key_pair.pem" ubuntu@namenode` *(atau)*
  - `ssh -i "key_pair.pem" ubuntu@ec2-xx-xx-xx`
- Buat public key dengan `ssh_keygen` di namenode dan kopi ke `authorized_keys` dengan perintah:
  - `ssh-keygen -f ~/.ssh/id_rsa -t rsa -P ""`
  - `~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys`
- Untuk koneksi password-less maka copy authorized_keys ke semua node dengan perintah:
  - `ssh datanode1 'cat >> ~/.ssh/authorized_keys' < ~/.ssh/id_rsa.pub`
  - `ssh datanode2 'cat >> ~/.ssh/authorized_keys' < ~/.ssh/id_rsa.pub`
  - `ssh datanode3 'cat >> ~/.ssh/authorized_keys' < ~/.ssh/id_rsa.pub`

Contoh Koneksi SSH ke Semua Node

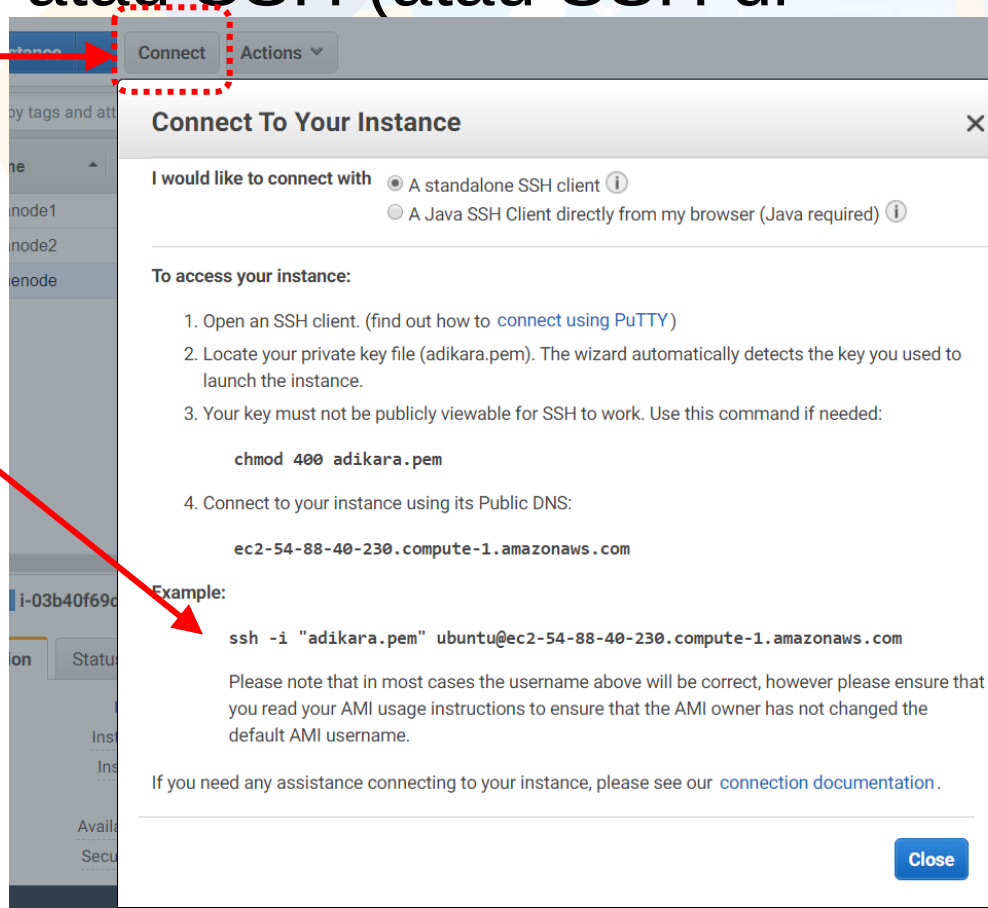
- Coba koneksi dari master/namenode ke datanode1 dan datanode2
  - `ssh datanode1`
  - `ssh datanode2`
- Begitu juga coba koneksi dari datanode1/ datanode2 ke namenode
  - `ssh namenode`



Konfigurasi NameNode dan DataNode

Menyiapkan Master (Namenode)

- Buat koneksi ke instance yang akan menjadi master melalui PuTTY atau SSH (atau SSH di Bash di WSL)
- Klik tombol Connect untuk mengetahui cara Connect
- Di terminal gunakan perintah ssh, misalnya:
  - `ssh -i "nama_ke.pem" ubuntu@ec2-xx-xx-xxx.amazonaws.com`



Menyiapkan Master (Namenode)

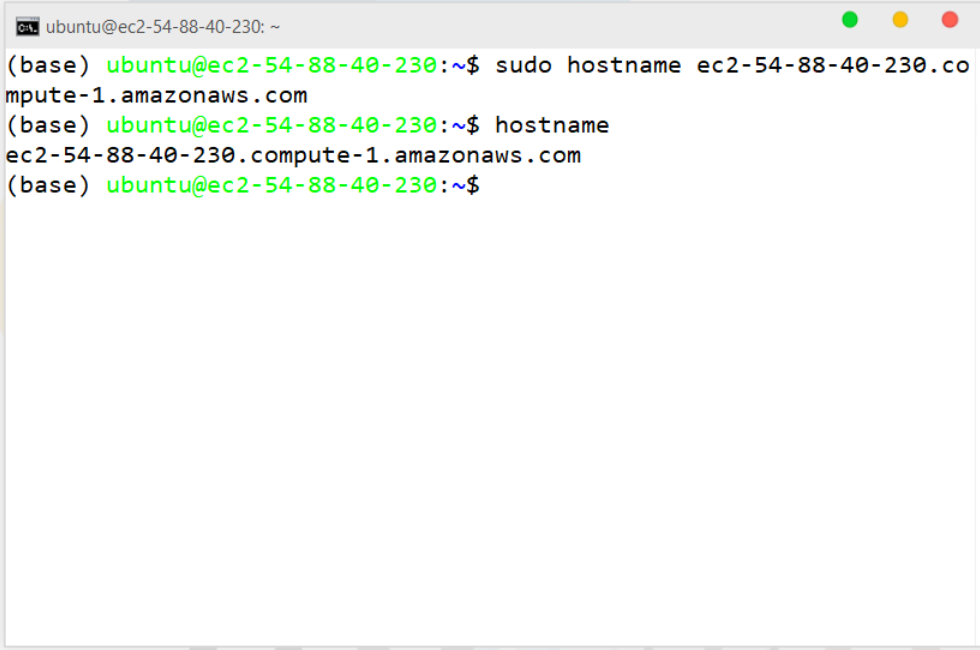
1. **Login ke ssh (slide sebelumnya)**
2. **Ubah nama hostname**
   - Nilai awalnya biasanya ip-xx-xx-xx-xx

- Ganti ke public DNS dengan perintah

  `sudo hostname <public_dns>`

- Misalnya

  `sudo hostname ec2-54-88-40-230.compute-1.amazonaws.com`



- Cek perubahan nama host dengan perintah

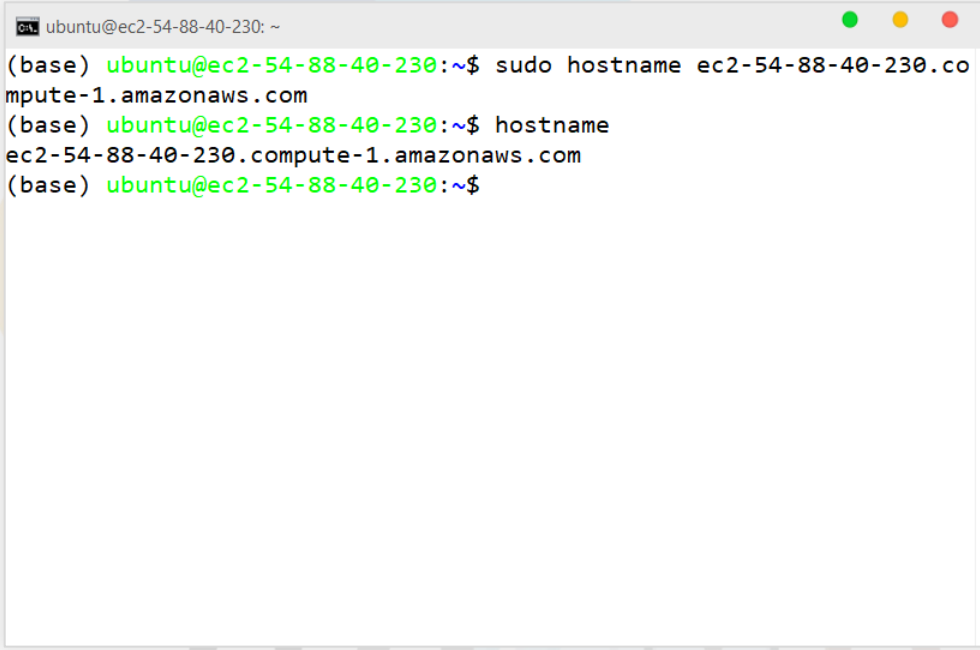
  `hostname`

Menyiapkan Slave (Datanode)

1. **Login ke datanode, mis**
   - `ssh datanode1`
2. **Ubah nama hostname untuk datanode**
   - Nilai awalnya biasanya ip-xx-xx-xx-xx
   - Ganti ke public DNS dengan perintah

     `sudo hostname <public_dns>`
   - Misalnya

     `sudo hostname ec2-54-88-40-230.compute-1.amazonaws.com`

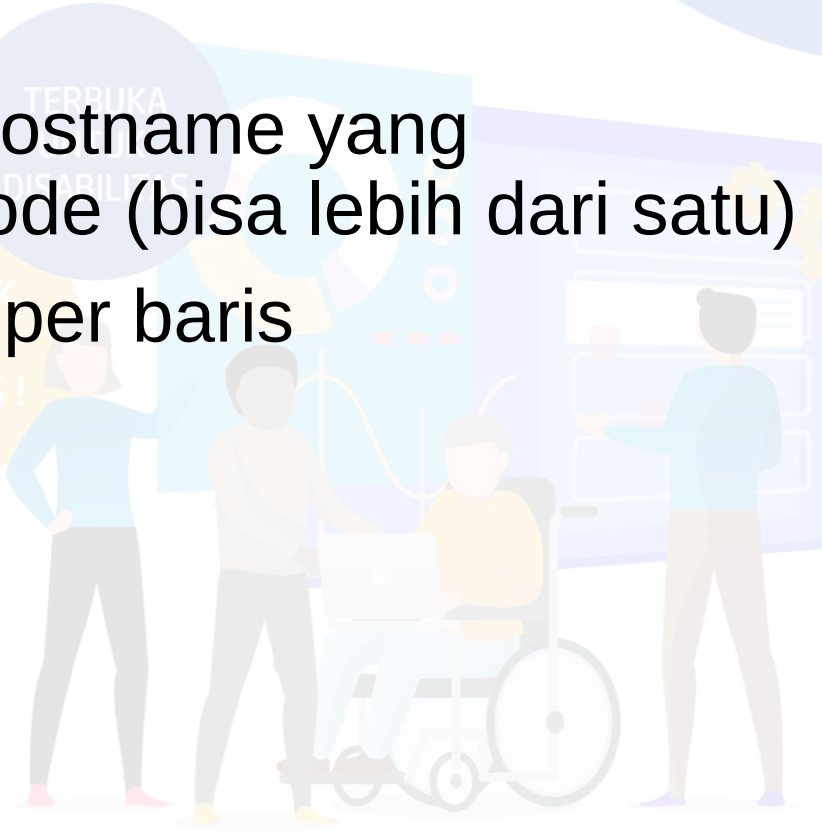


- Cek perubahan nama host dengan perintah

  `hostname`

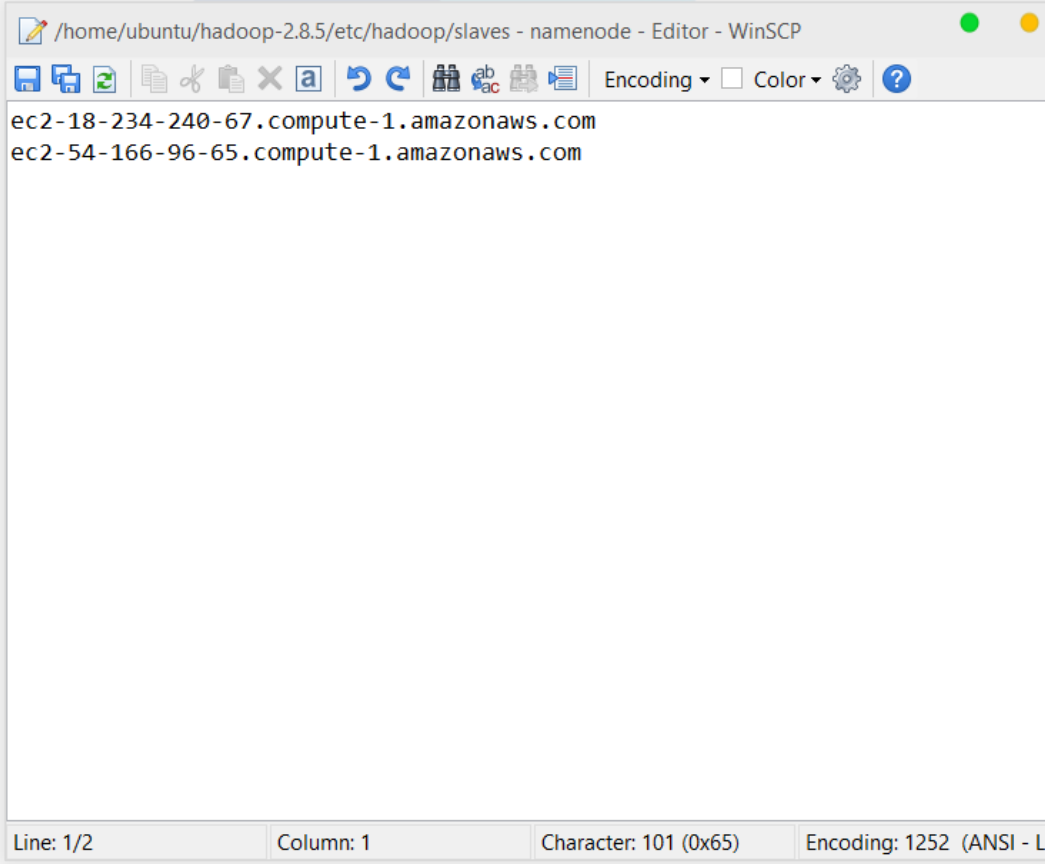
Konfigurasi Fail "masters"

- File ini berisi nama-nama hostname yang digunakan sebagai namenode (bisa lebih dari satu)
- Nama hostname dituliskan per baris



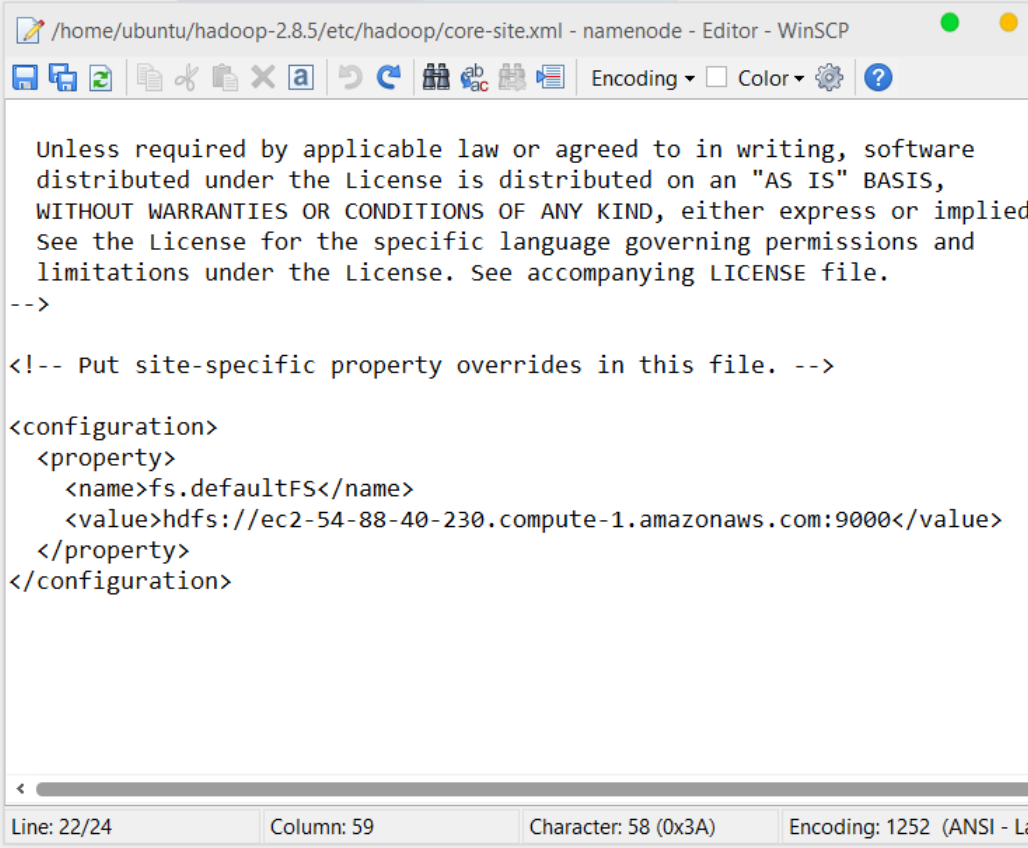
Konfigurasi Fail "slaves"

- File ini berisi nama-nama hostname yang digunakan sebagai datanode (bisa lebih dari satu)
- Nama hostname dituliskan per baris



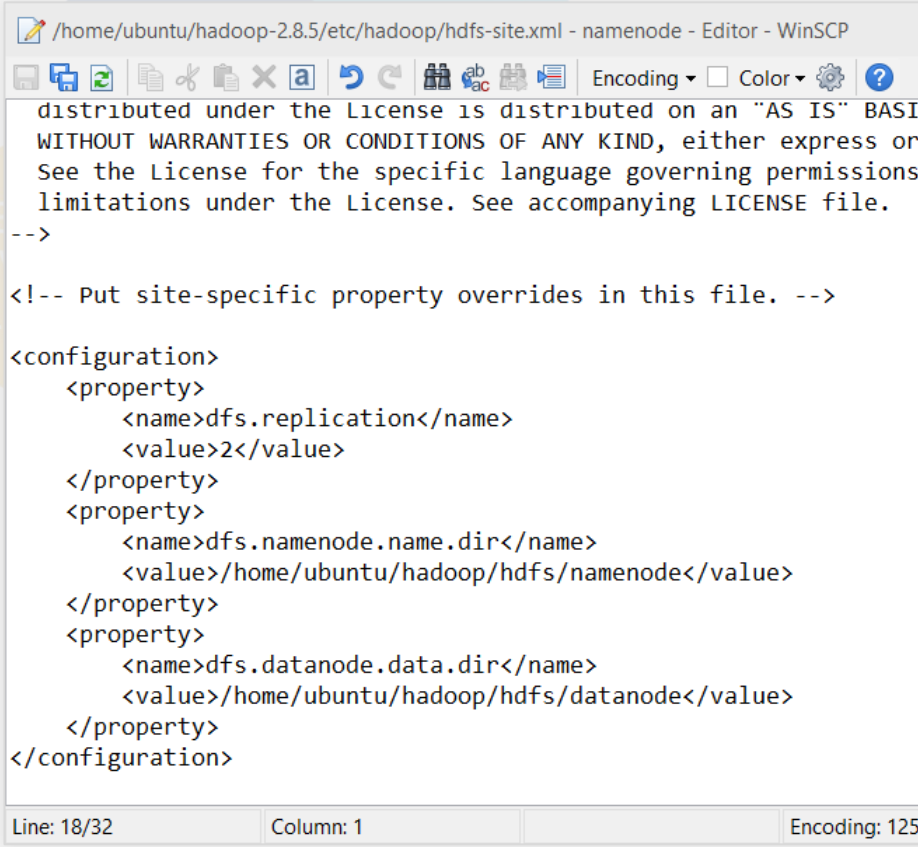
Konfigurasi Fail "core-site.xml"

- Ganti localhost pada <value> dengan public DNS dari namenode



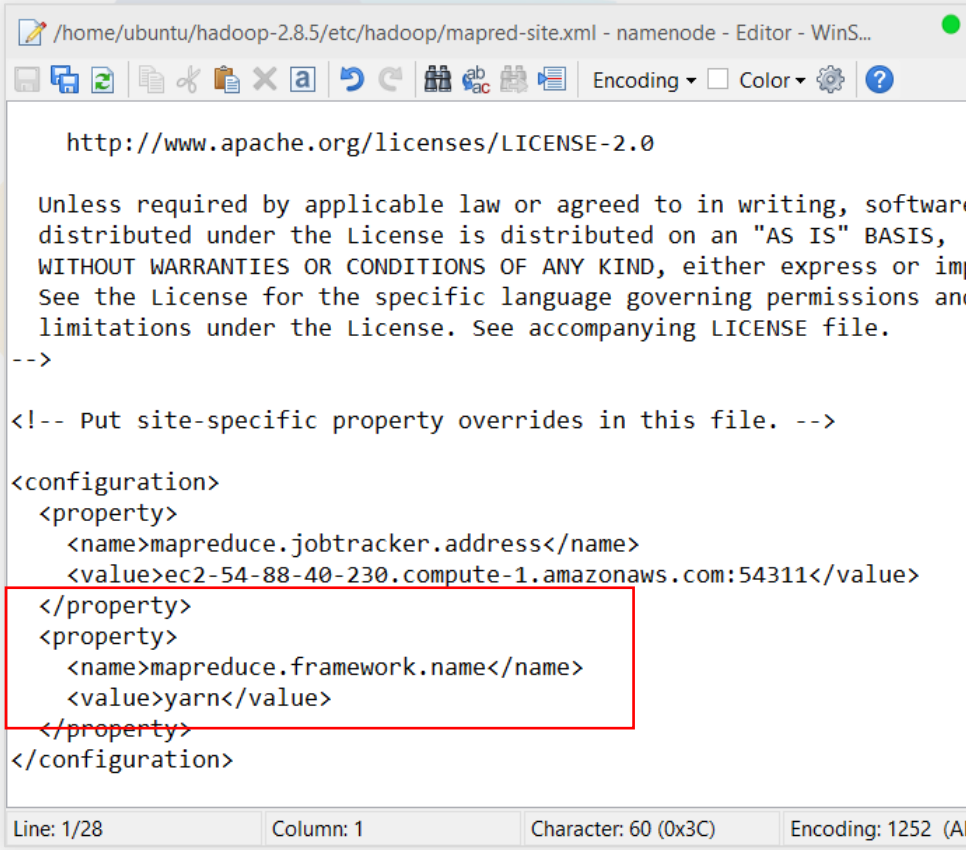
Konfigurasi Fail "hdfs-site.xml"

- Pada property `dfs.replication`, isi <value> sebanyak datanode yang digunakan
  - contoh di sini ada 2, datanode1 dan datanode2
- Pada `dfs.namenode.name.dir`, isi <value> dengan lokasi path nantinya metadata dari namenode disimpan
- Pada `dfs.datanode.data.dir`, isi <value> dengan lokasi path nantinya data sebenarnya dari datanode disimpan
- Lokasi ini harus ada (buat dulu) dan mendapat permission untuk read write



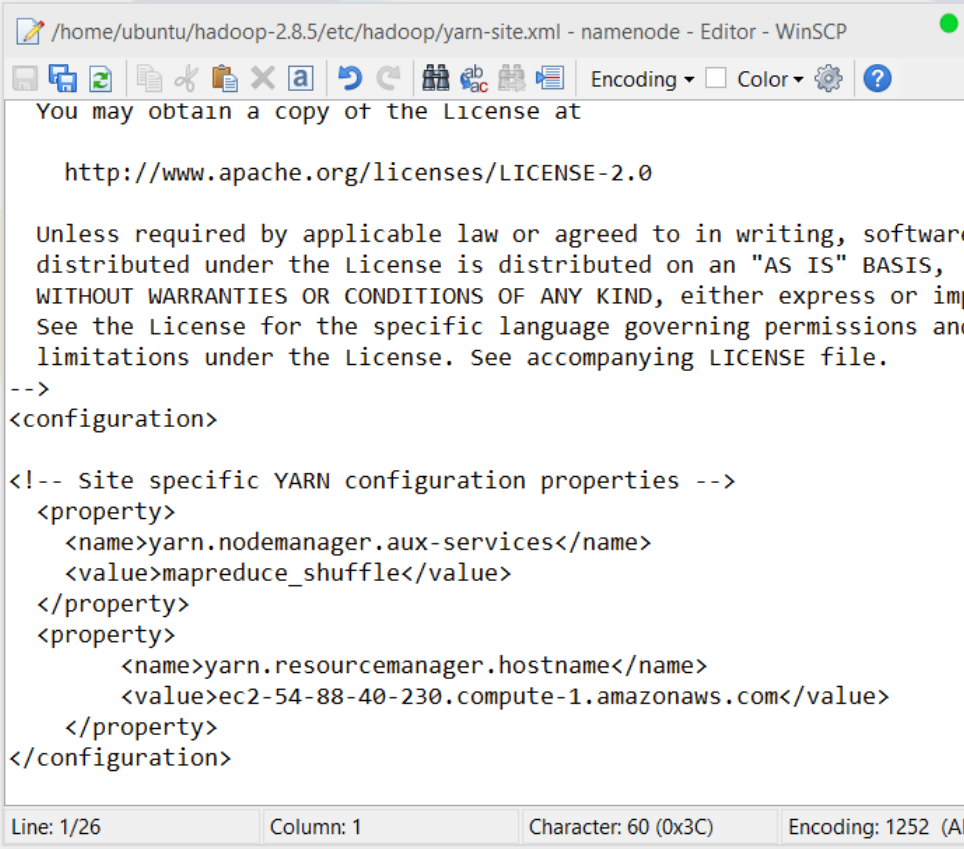
Konfigurasi Fail "mapred-site.xml"

- Untuk Hadoop 1,isi value pada `mapreduce.jobtracker.address`, dengan public DNS dari namenode dan port default 54311
- Untuk > Hadoop 2 tidak perlu diisikan, yang penting mengisi `mapreduce.framework.name` dengan `yarn`
  - Tugas JobTracker digantikan oleh YARN
  - Sebagai gantinya isi di `yarn.resourcemanager.address` di `yarn-site.xml`



Konfigurasi Fail "yarn-site.xml"

- Buat dan isi property `yarn.resourcemanager.hostname` dengan "public DNS" saja dari namenode
  - dengan port default
- *atau*
- Buat dan isi property `yarn.resourcemanager.address` dengan "public DNS:port"



Sudo install SCP

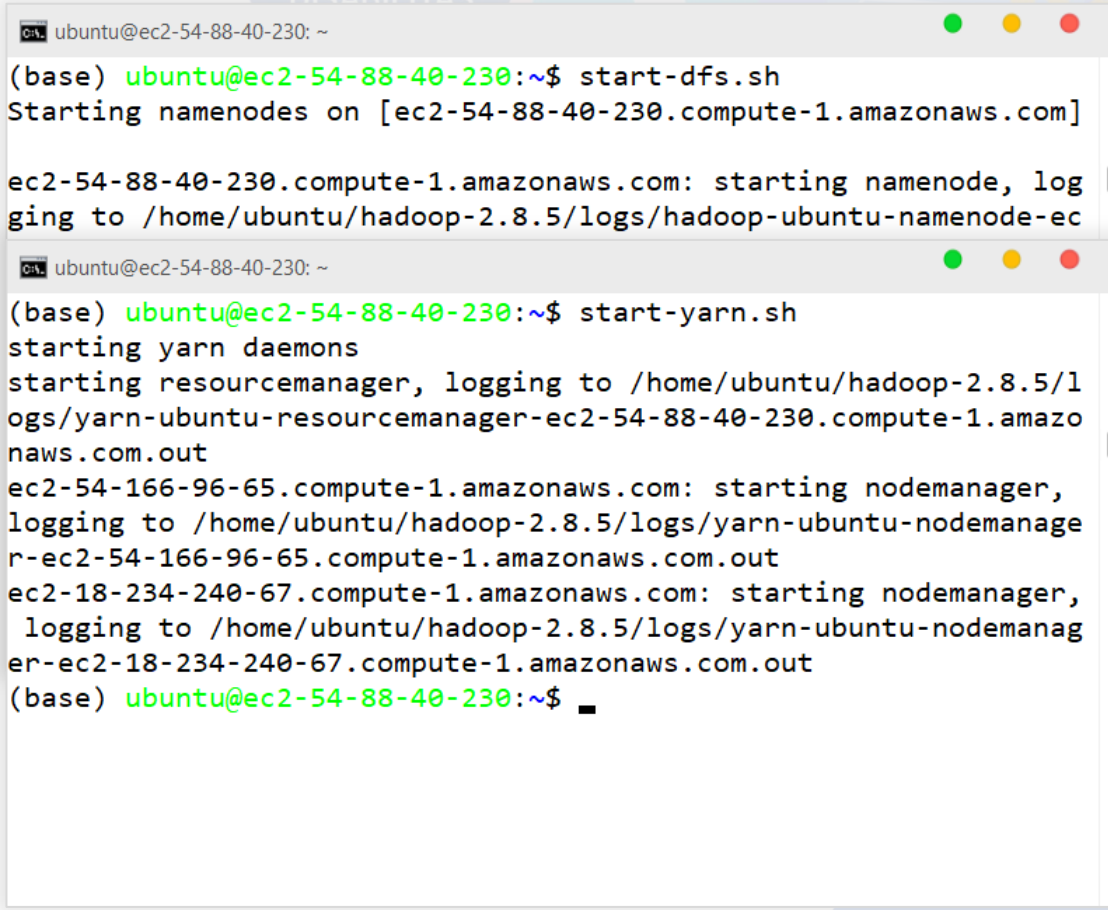
Kopikan Fail Konfigurasi ke Datanodes

- Melalui PuTTY atau SSH, pindah current directory ke lokasi konfigurasi Hadoop
  - Contoh lokasi: /home/ubuntu/hadoop-2.8.5/etc/hadoop/
  - cd /home/ubuntu/hadoop-2.8.5/etc/hadoop/
- Kopikan konfigurasi Namenode ke semua Datanode
  - scp *.xml datanode1:/home/ubuntu/hadoop-2.8.5/etc/hadoop/
  - scp *.xml datanode2:/home/ubuntu/hadoop-2.8.5/etc/hadoop/

Menjalankan Hadoop

Menjalankan Hadoop

- Masuk ke folder `hadoop/sbin`
  - Kecuali bila sudah masuk di PATH di ~/.bashrc, maka bisa panggil perintah dari path/folder apa saja
- Ketikkan perintah
  - `start-dfs.sh`
  - `start-yarn.sh`
  - *atau* `start-all.sh` (akan deprecated/ dihapus di versi baru)
- Bila gagal cek di fail log



Cek Proses Hadoop yang Berjalan

- Apabila Hadoop berhasil berjalan di namenode dan datanode maka akan muncul beberapa proses.
- Login ke namenode dan datanode lewat PuTTY/SSH
- Cek dengan `jps`
  - Namenode

```
ubuntu@ec2-54-88-40-230: ~
(base) ubuntu@ec2-54-88-40-230:~$ jps
1905 SecondaryNameNode
1682 NameNode
2323 Jps
2060 ResourceManager
(base) ubuntu@ec2-54-88-40-230:~$
```

  - Datanode

```
ubuntu@ec2-18-234-240-67: ~
(base) ubuntu@ec2-18-234-240-67:~$ jps
1536 DataNode
1680 NodeManager
2210 Jps
(base) ubuntu@ec2-18-234-240-67:~$
```

Menghentikan Hadoop

Menghentikan Proses Hadoop

- Masuk ke folder `hadoop/sbin`
  - Kecuali bila sudah masuk di PATH di ~/.bashrc, maka bisa panggil perintah dari path/folder apa saja
- Ketikkan perintah
  - `stop-dfs.sh`
  - `stop-yarn.sh`
  - *atau* `stop-all.sh` (akan deprecated/dihapus di versi baru)
- Cek dengan `jps`
  - Bila proses diberhentikan semua, maka tinggal proses `jps` saja yang muncul

IKUTI KAMI

DIGITAL TALENT SCHOLARSHIP

- digitalent.kominfo
- digitalent.kominfo
- DTS_kominfo
- Digital Talent Scholarship 2019

Pusat Pengembangan Profesi dan Sertifikasi
Badan Penelitian dan Pengembangan SDM
Kementerian Komunikasi dan Informatika
Jl. Medan Merdeka Barat No. 9
(Gd. Belakang Lt. 4 - 5)
Jakarta Pusat, 10110