**LAPORAN PRAKTIKUM**

**PRAKTIKUM INFRASTRUKTUR BIG DATA**

**PERTEMUAN -4**

****

Disusun oleh :

**Nama : Untung Leo**

**Nim : 215410022**

**Kelas : Informatika -1**

**Jenjang : S1**

**FAKULTAS TEKNOLOGI INFORMASI**

**UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA**

**YOGYAKARTA**

**2022/2023**

**Modul 4**

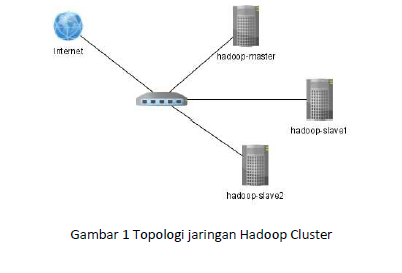
**Setup Hadoop Cluster**

**Dasar Teori**

Apache Hadoop adalah kumpulan utilitas perangkat lunak open-source yang memungkinkan distribusi kumpulan data dalam jumlah besar di seluruh kluster komputer menggunakan model pemrograman sederhana.

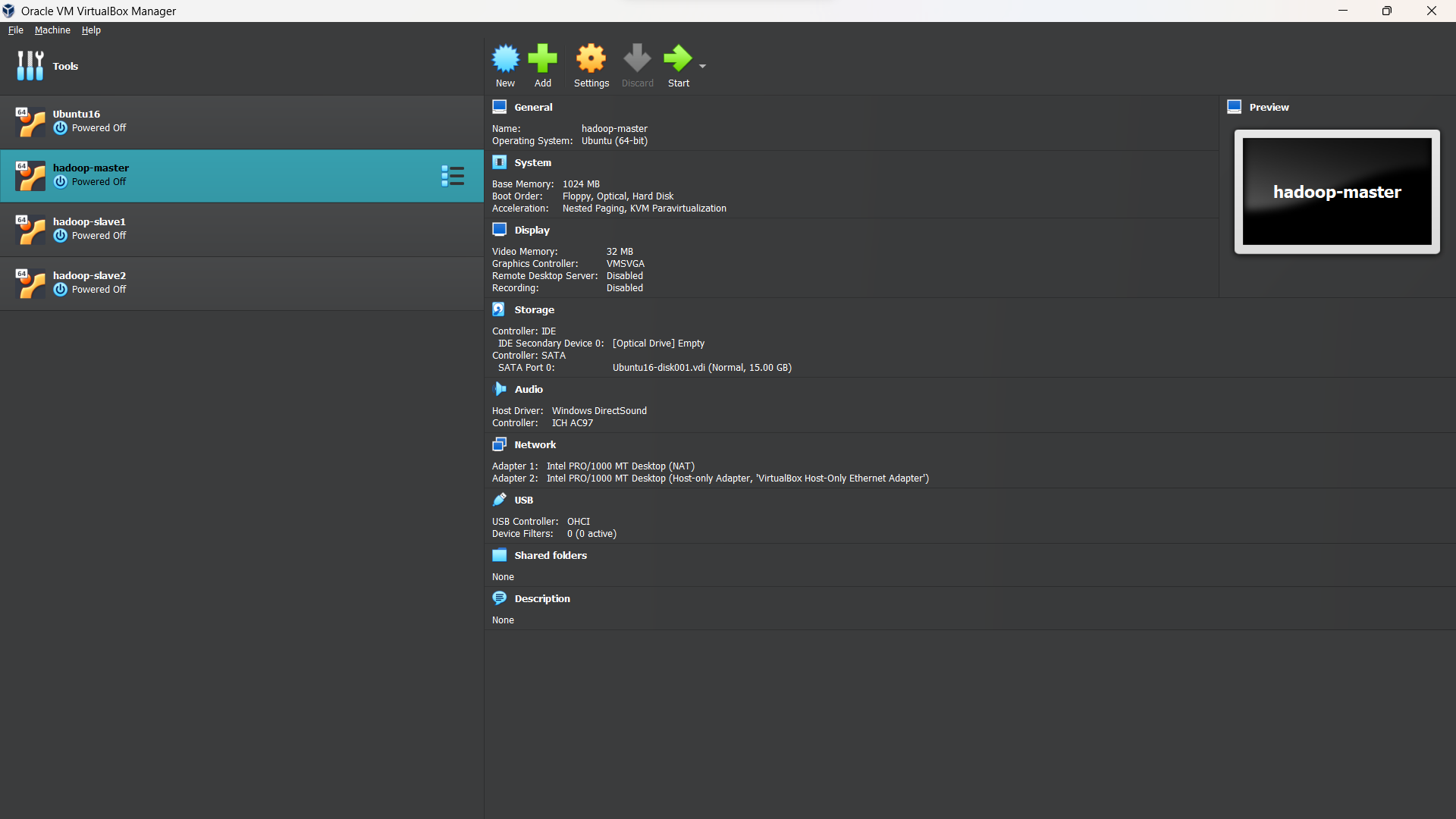
Hadoop menyediakan software framework untuk penyimpanan dan pemrosesan yang terdistribusi dari Big Data menggunakan MapReduce. Hadoop didesign untuk computer clusters (“adalah sekumpulan komputer yang terhubung dan saling bekerja bersama, sehingga seolah-olah seperti satu sistem”) komuditas hardware, Hadoop juga bisa digunakan pada cluster dengan higher-end hardware.

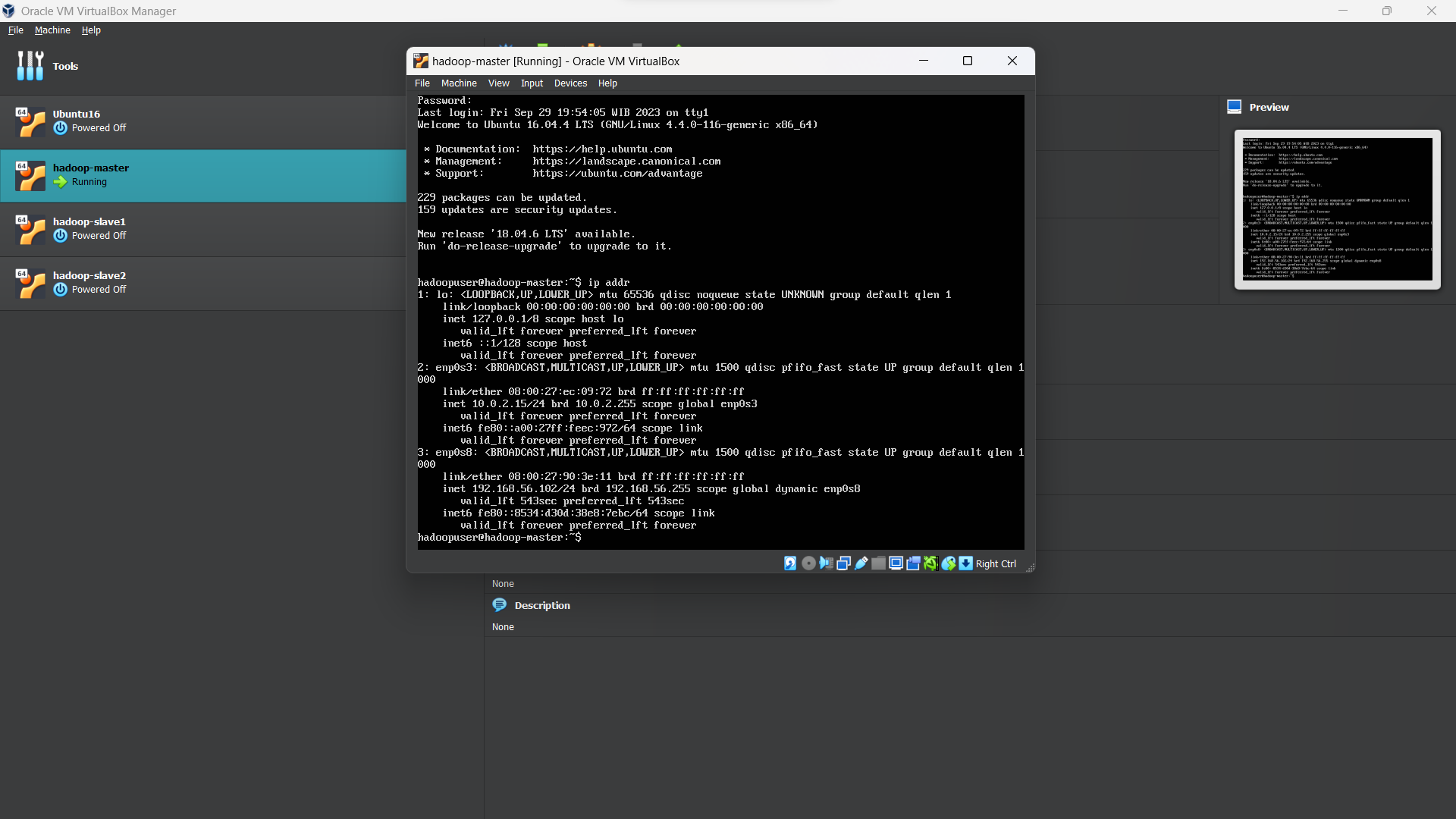
Pada praktikum ini kita akan mengimplementasikan hadoop cluster yang terdiri dari 3 node yaitu, hadoop-master, hadoop-slave1, hadoop-slave2. Dengan topologi jaringan seperti pada gambar 1. Dalam praktikum ini, kita perlu mempersiapkan environment terlebih dahulu sebelum memulai mendeploy hadoop cluster.

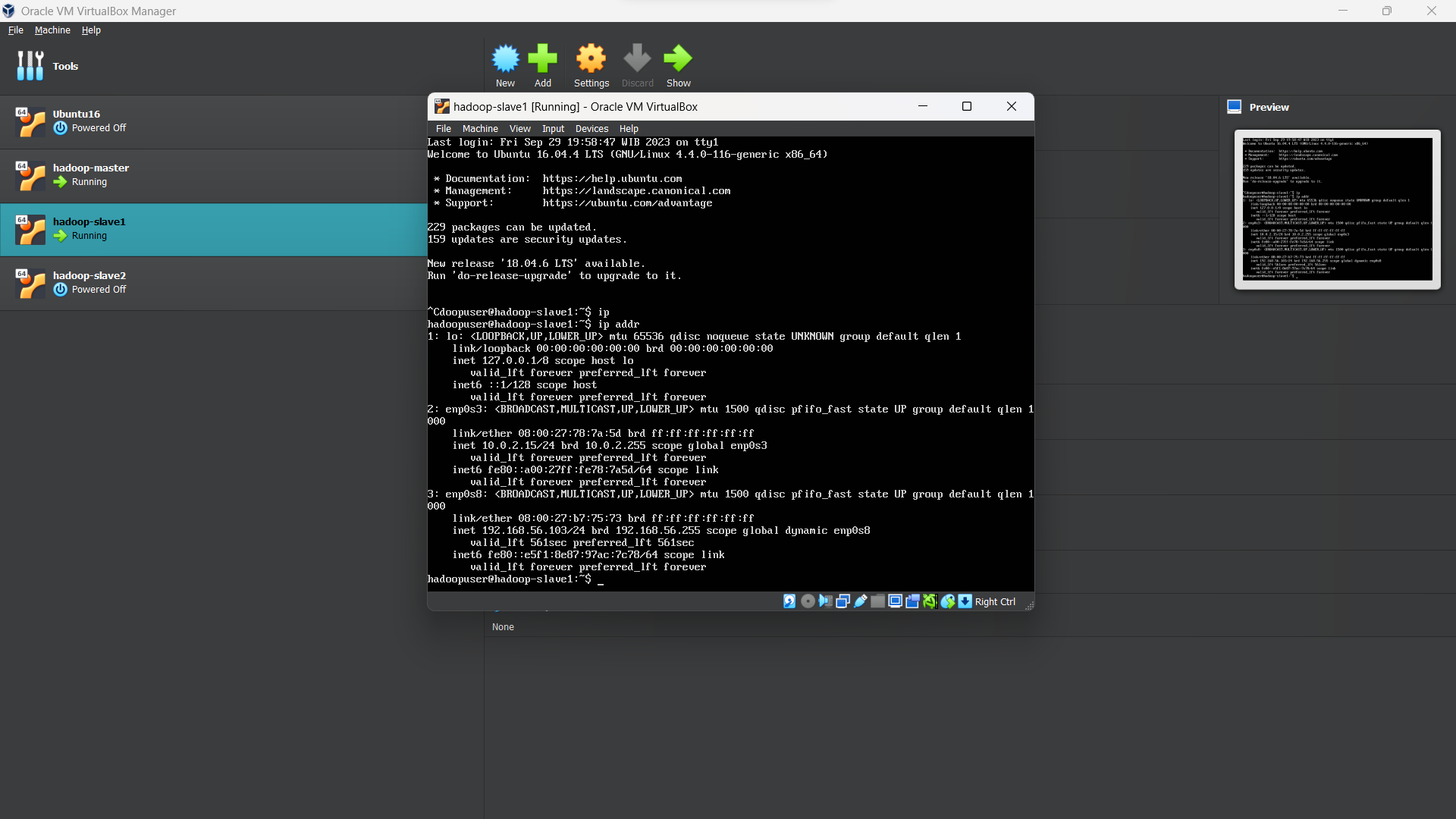


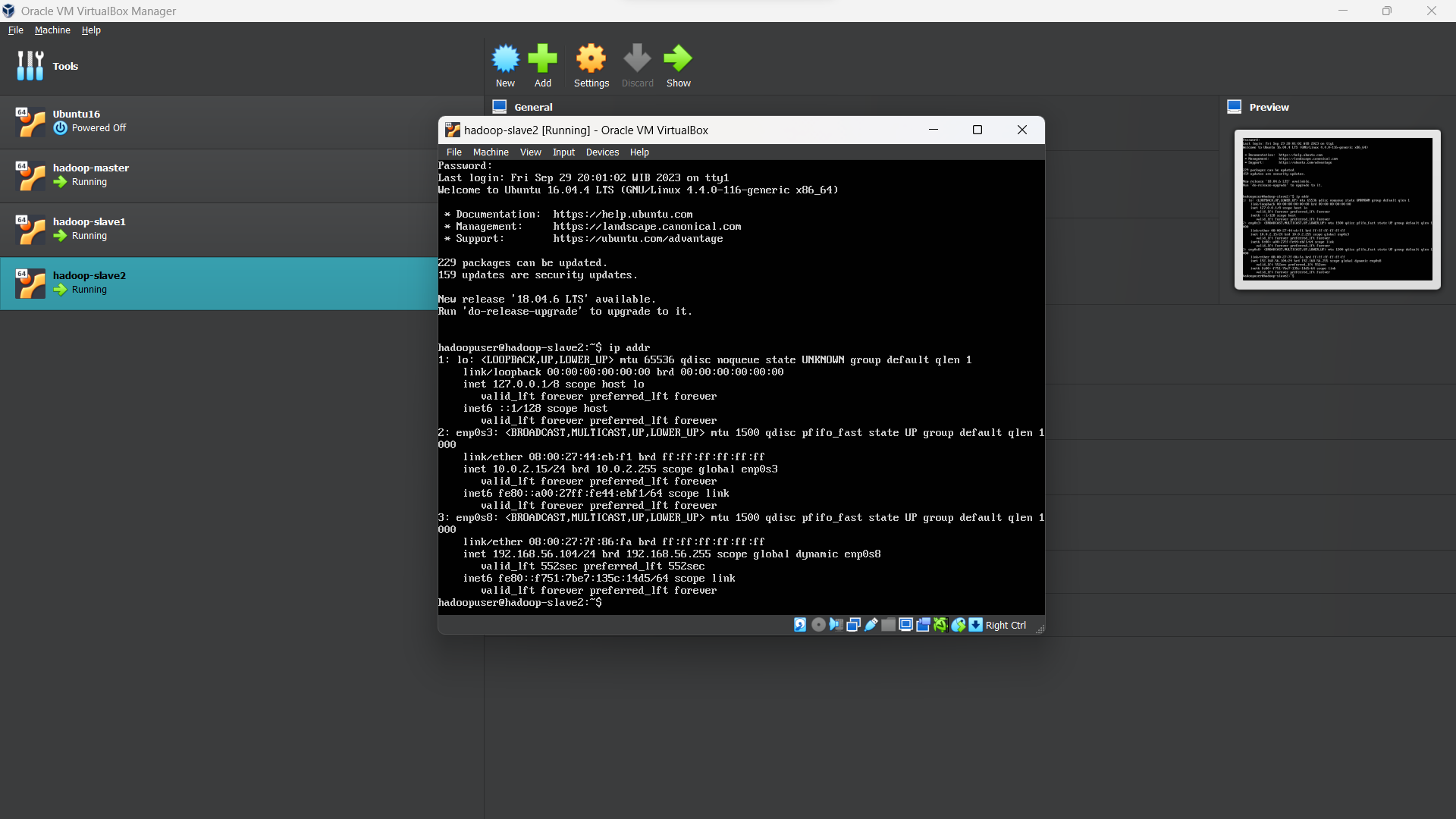
Langkah-langkah praktikum

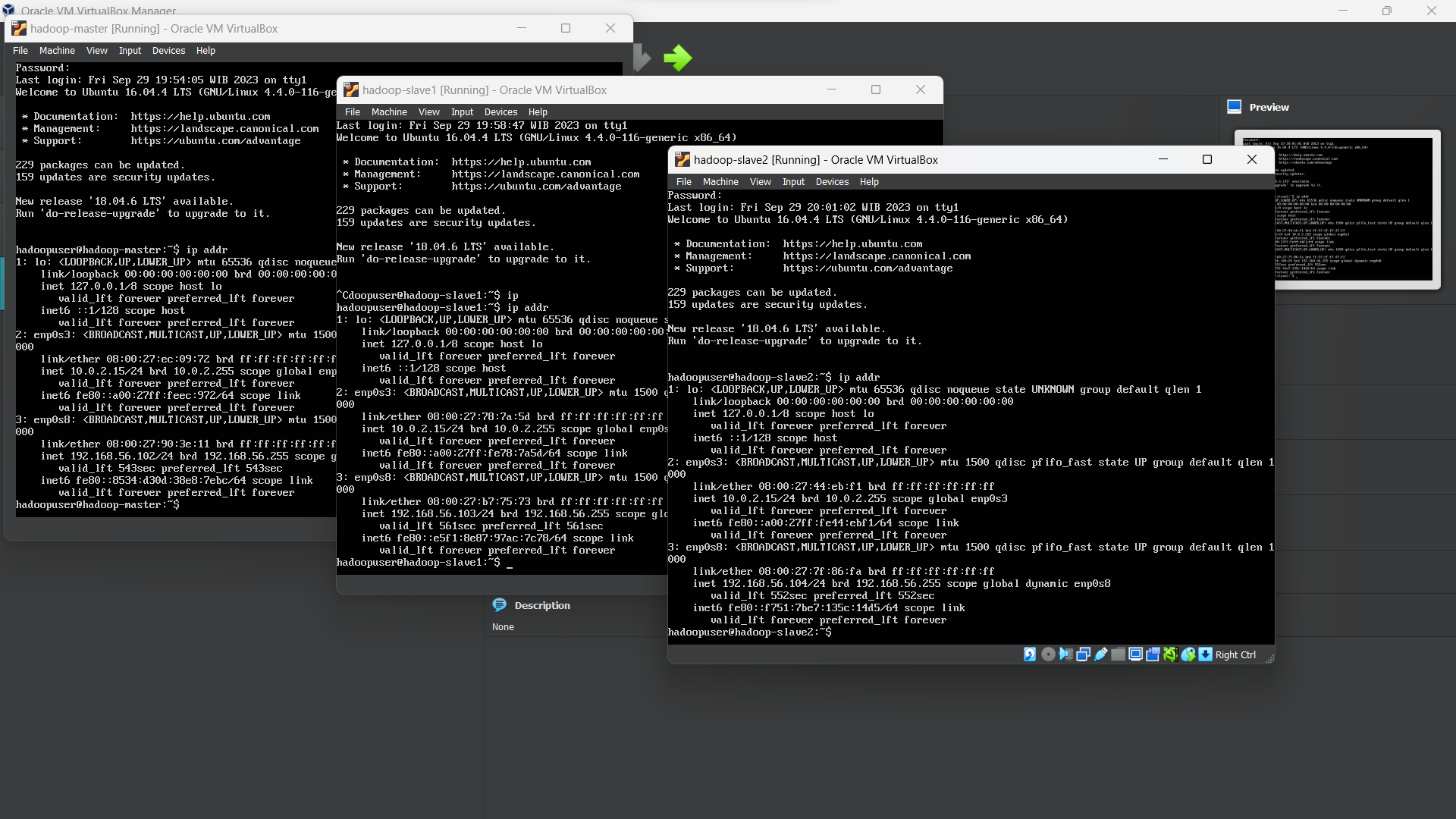
1. Jalankan semua node **(hadoop-master, hadoop-slave1, & hadoop-slave2)** pada virtualbox yang telah dibuat pada praktikum sebelumnya, Login dangan menggunakan user: **hadoopuser**, password: **hadoopuser**, dan cek IP address dari masing-masing node dengan menggunakan perintah “**ip addr**”. Kemudian ping dari satu node ke node yang lain untuk memastikan semua node terkoneksi dalam satu jaringan.





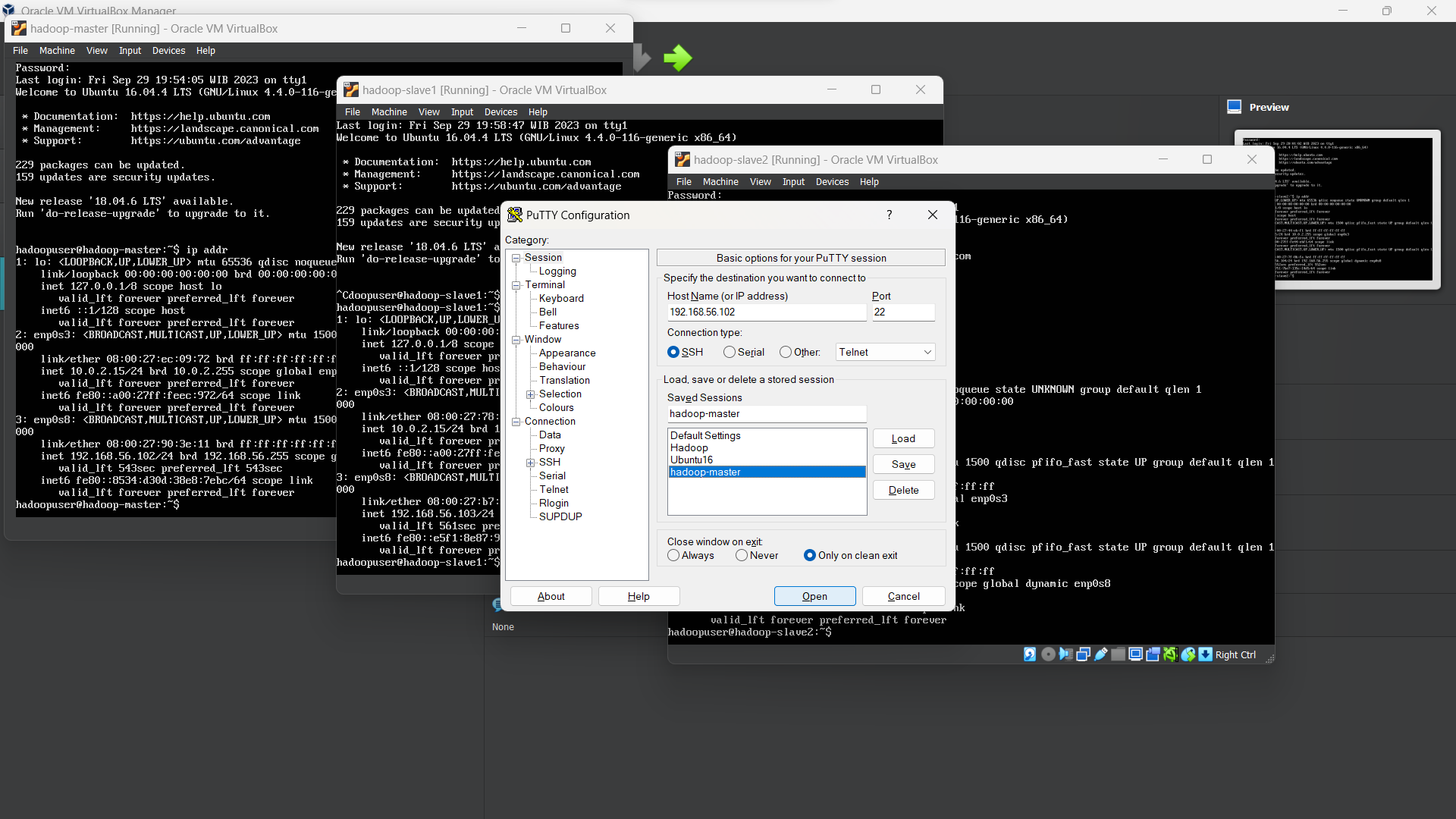


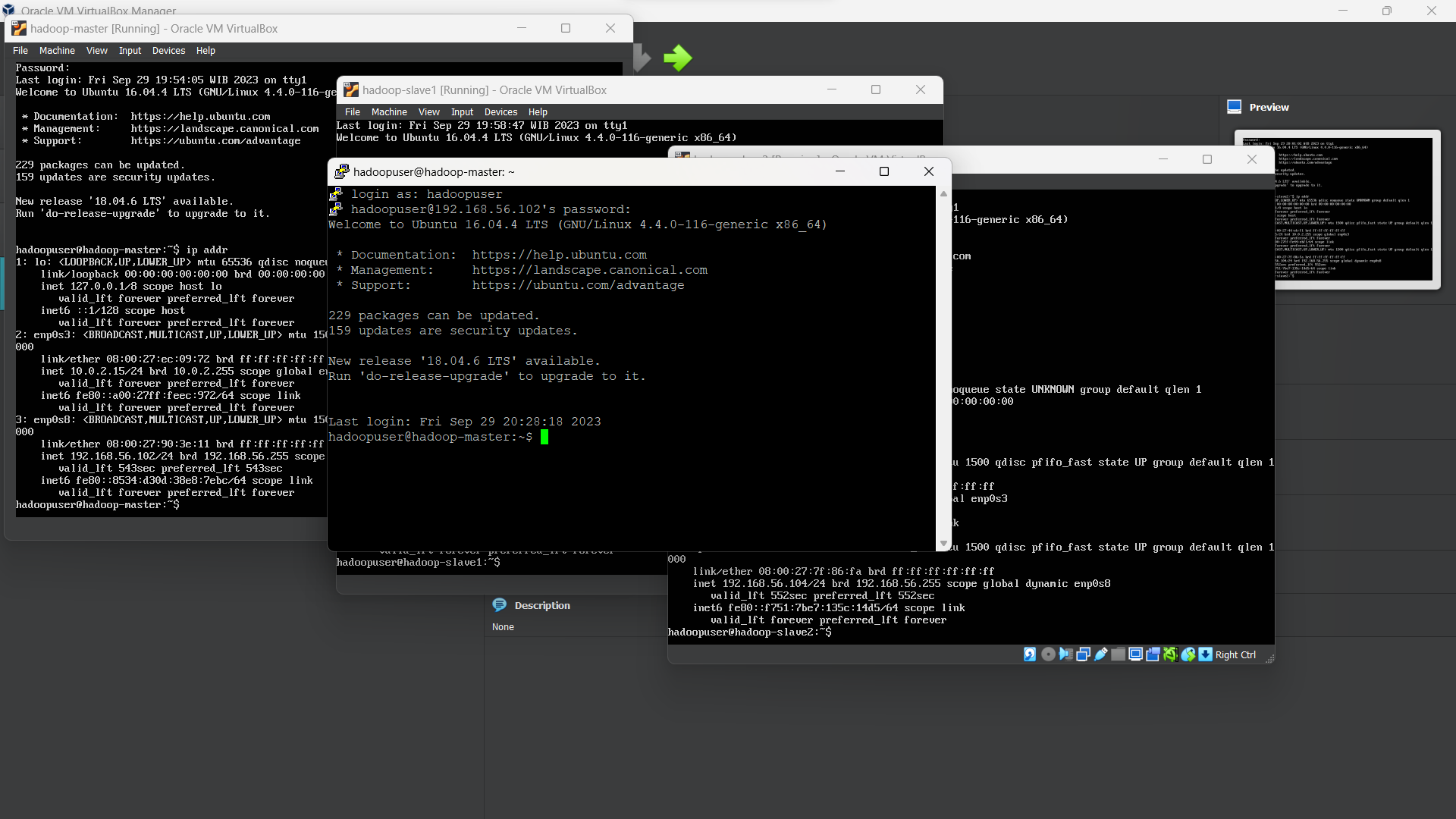




Dilangkah yang pertama ini kita di minta menjalankan semua virual machin yang sudah kita cloning di pertemuang sebelumnya.

1. Pada hadoop-master, login via **putty** dengan menggunakan user: **hadoopuser**, password: **hadoopuser**.

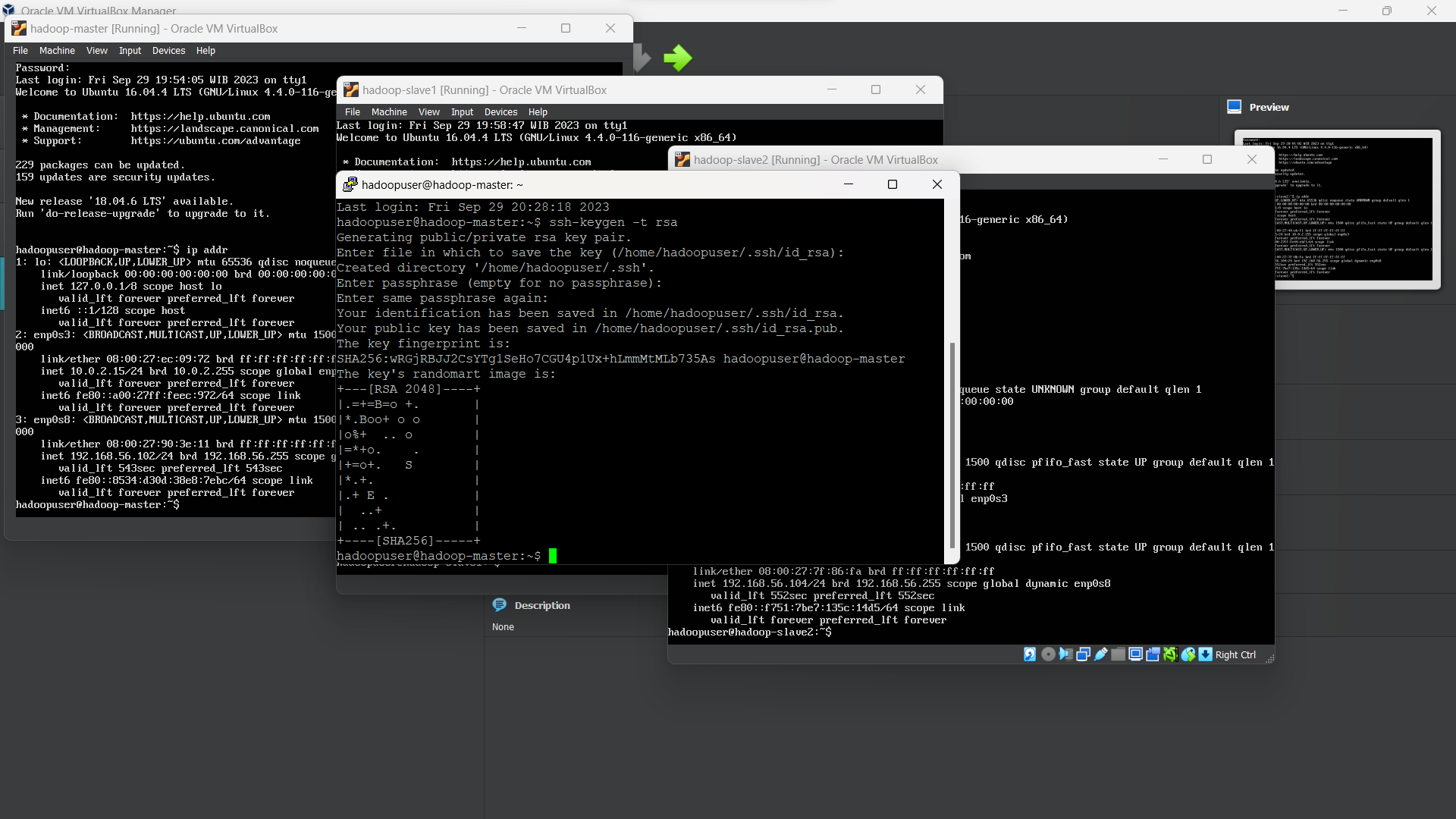




Di langkah yang kedua ini kita di minta login melalui putty dengan menggunakan ip addres hadoop-master kemudian kita login menggunakan user : hadoopuser dan password : hadoopuser sehingga muncul tampilan seperti pada gambar di atas.

1. Pada **hadoop-master**, Buat ssh key, kemudian copy ke masing-masing node. Jika muncul notifikasi untuk mengetikkan **passphrase** maka tekan ENTER untuk mengkosongi atau melewatinya.

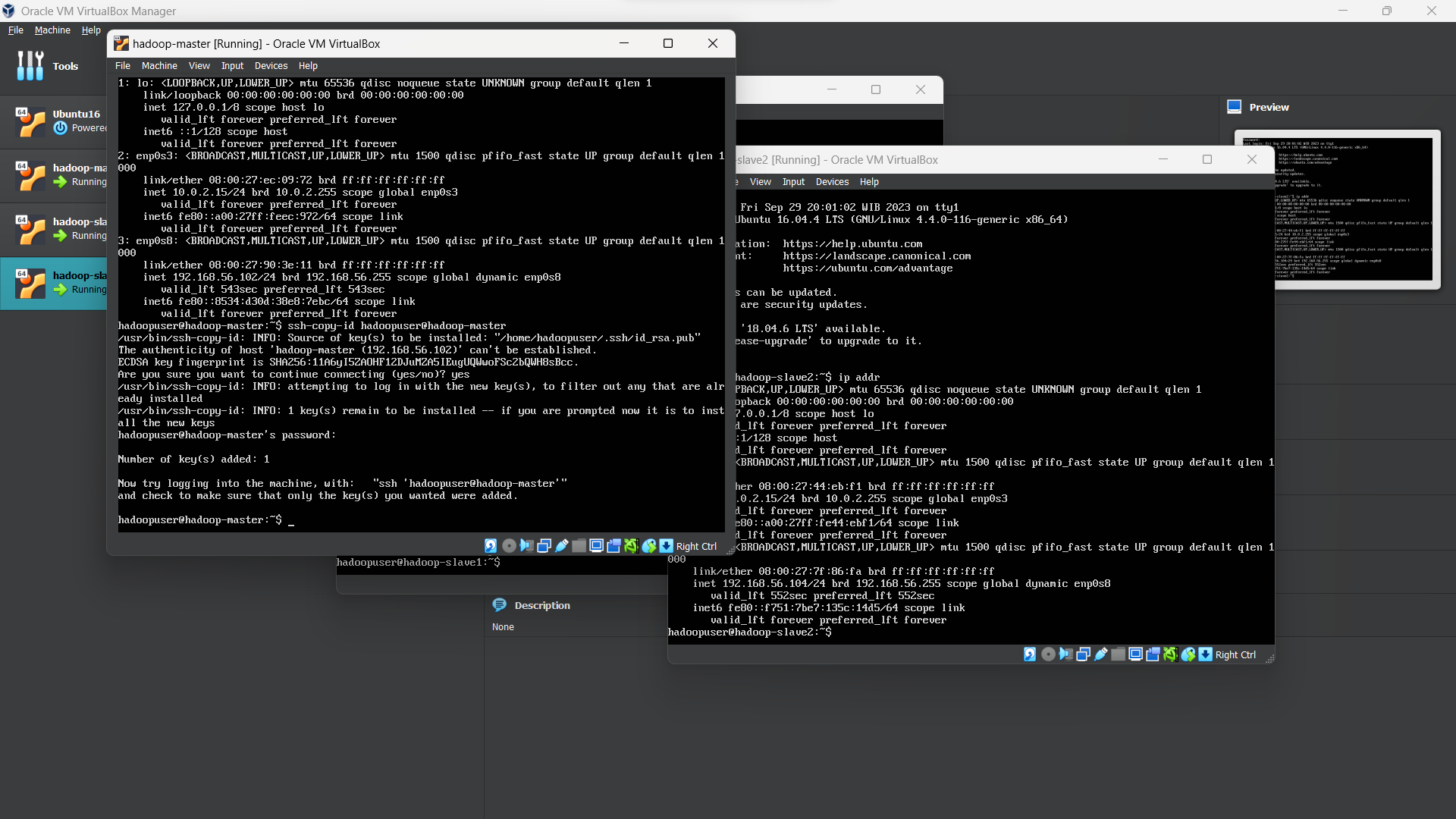
$ ssh-keygen -t rsa

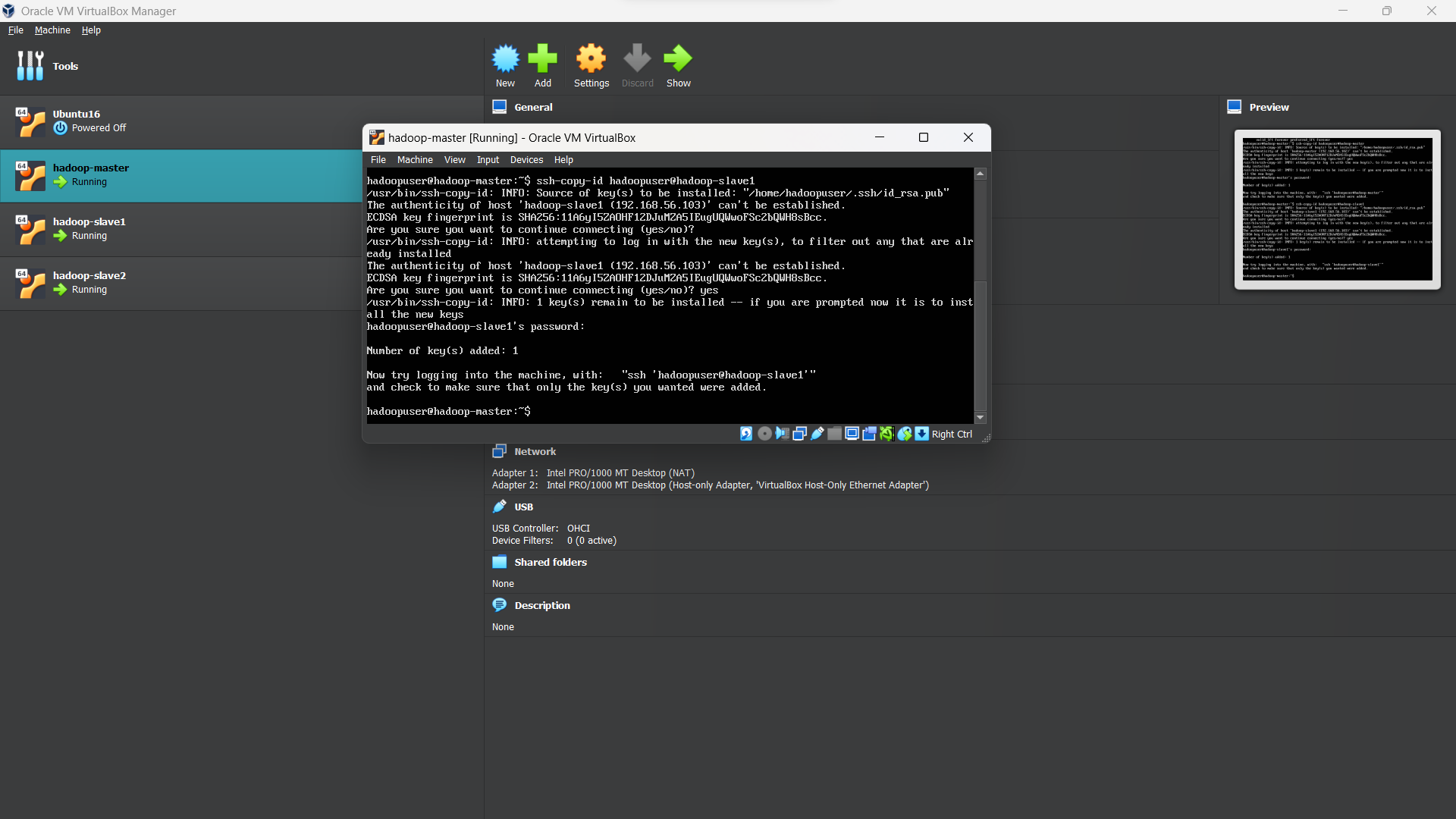


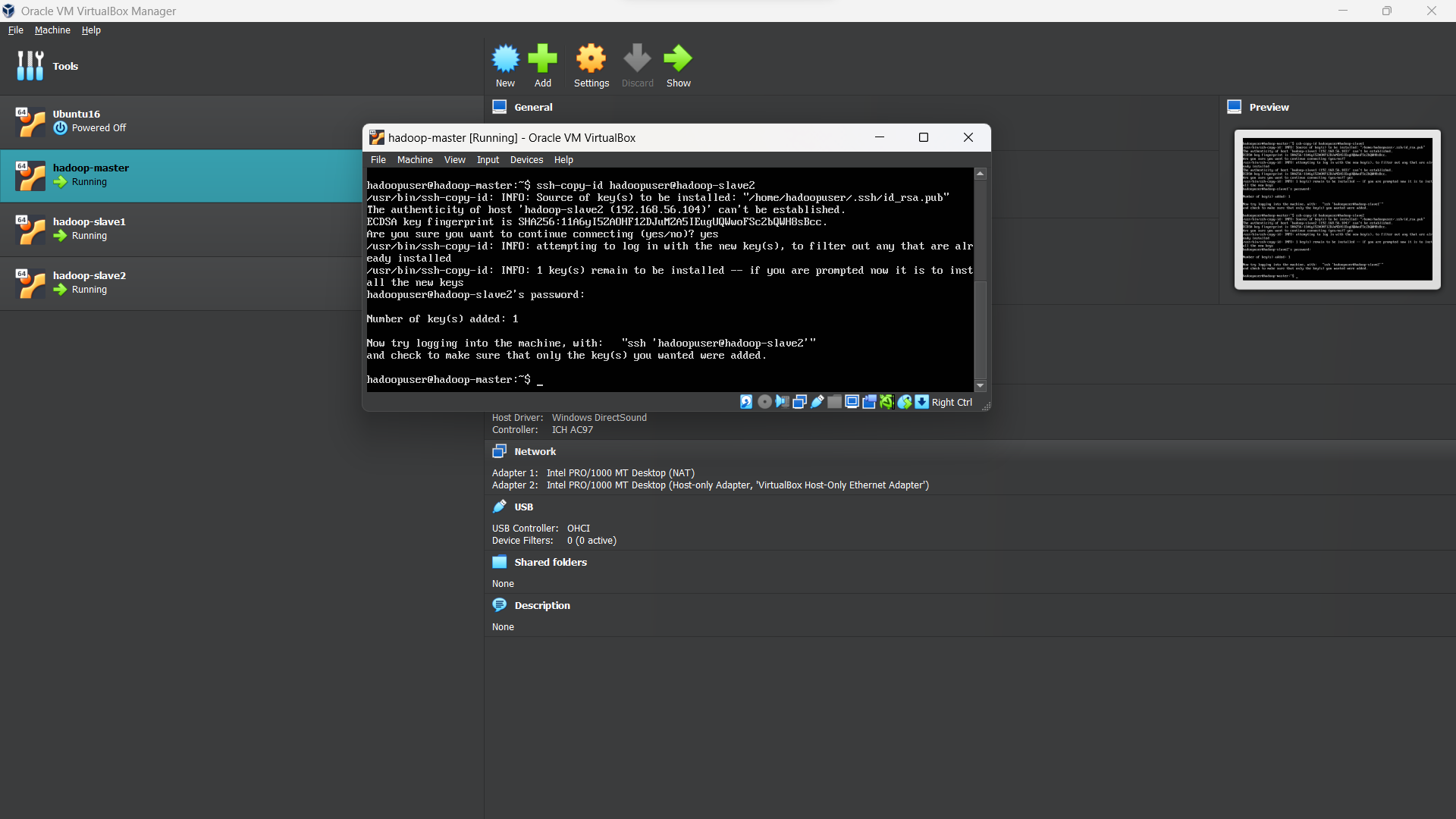
Disini kita akan membuat ssh key untuk menghasilkan pasangan kunci SSH dengan algoritma enkripsi RSA. Ini adalah langkah pertama dalam mengatur otentikasi SSH berbasis kunci, yang lebih aman daripada menggunakan kata sandi. kemudian copy ke masing-masing node. Jika muncul notifikasi untuk mengetikkan passphrase maka tekan ENTER untuk mengkosongi atau melewatinya.

1. Dari **hadoop-master**, Kemudian copy ssh key ke masing-masing node. Jalankan perintah berikut pada **semua node Virtual Machine** (hadoop-master, hadoop-slave1, & hadoop-slave2)

$ ssh-copy-id hadoopuser@hadoop-master$ ssh-copy-id hadoopuser@hadoop-slave1$ ssh-copy-id hadoopuser@hadoop-slave2







Setelah kita membuat ssh key sekarang kita akan membuat lagi di virtual machine yang lain yang telah kita cloning di pertemuan sebelumnya. Seperti yang terlihat pada gambar di atas dengan menggunakan perintah yang tersediah di atas.

1. Pada **hadoop-master**, buka file **core-site.xml** untuk dikonfigurasi.

$ sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml

<configuration>

<property>

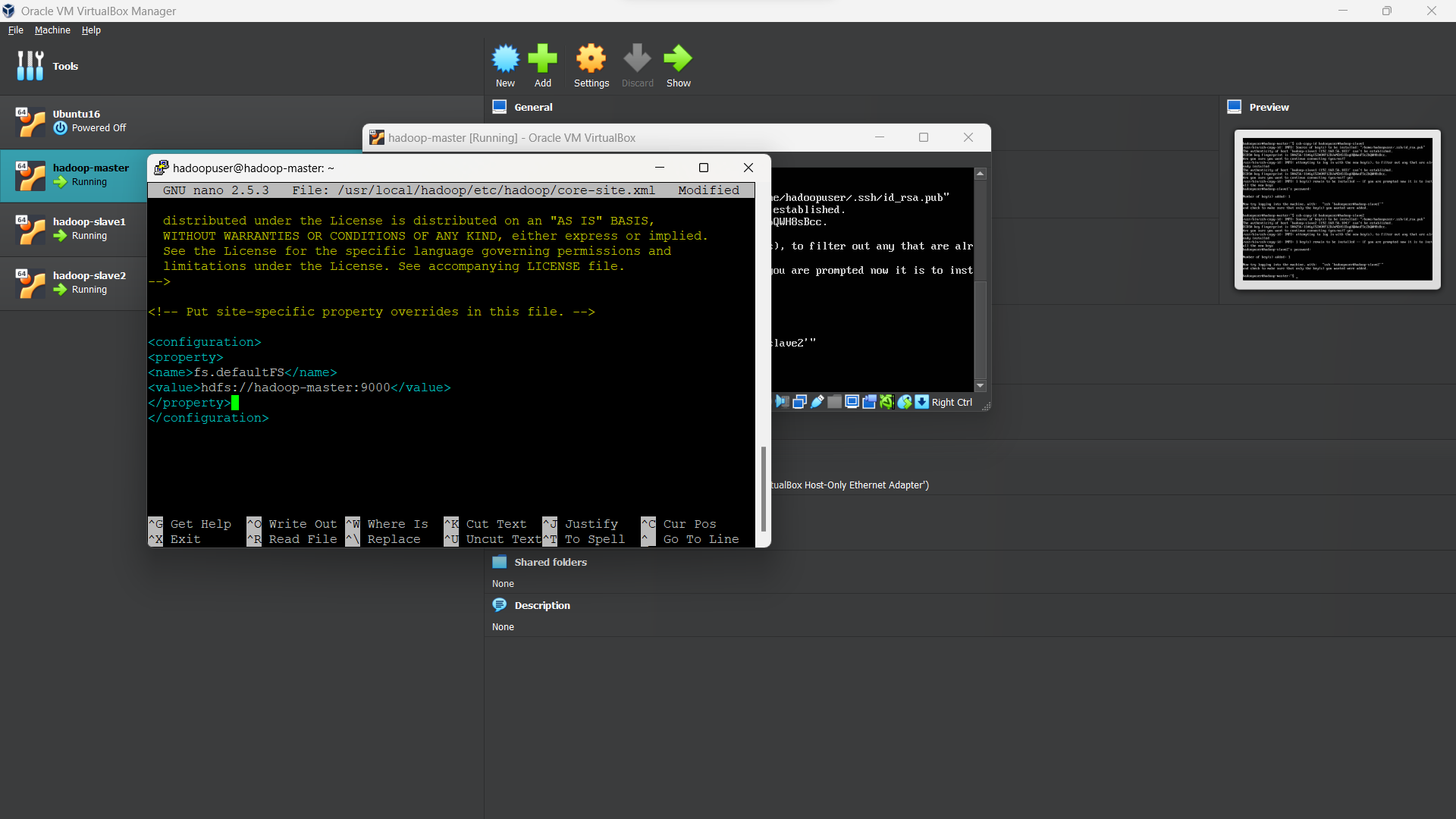
<name>fs.defaultFS</name>

<value>hdfs://hadoop-master:9000</value>

</property>

</configuration>

Untuk keluar (menyimpan) dari editor Tekan “ctrl+X, Y, Enter”



Disini kita akan buka file core-site.xml untuk melakukan kofigurasi di dalam machine nya sendiri untuk tempat kita menaruh file configurasinya ada di paling bawah seperti pada gambar di atas.

1. Pada hadoop-master, buka file hdfs-site.xml untuk dikonfigurasi.

$ sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml

<configuration>

<property>

<name>dfs.namenode.name.dir</name><value>/usr/local/hadoop/data/nameN

ode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name><value>/usr/local/hadoop/data/dataN

ode</value>

</property>

<property>

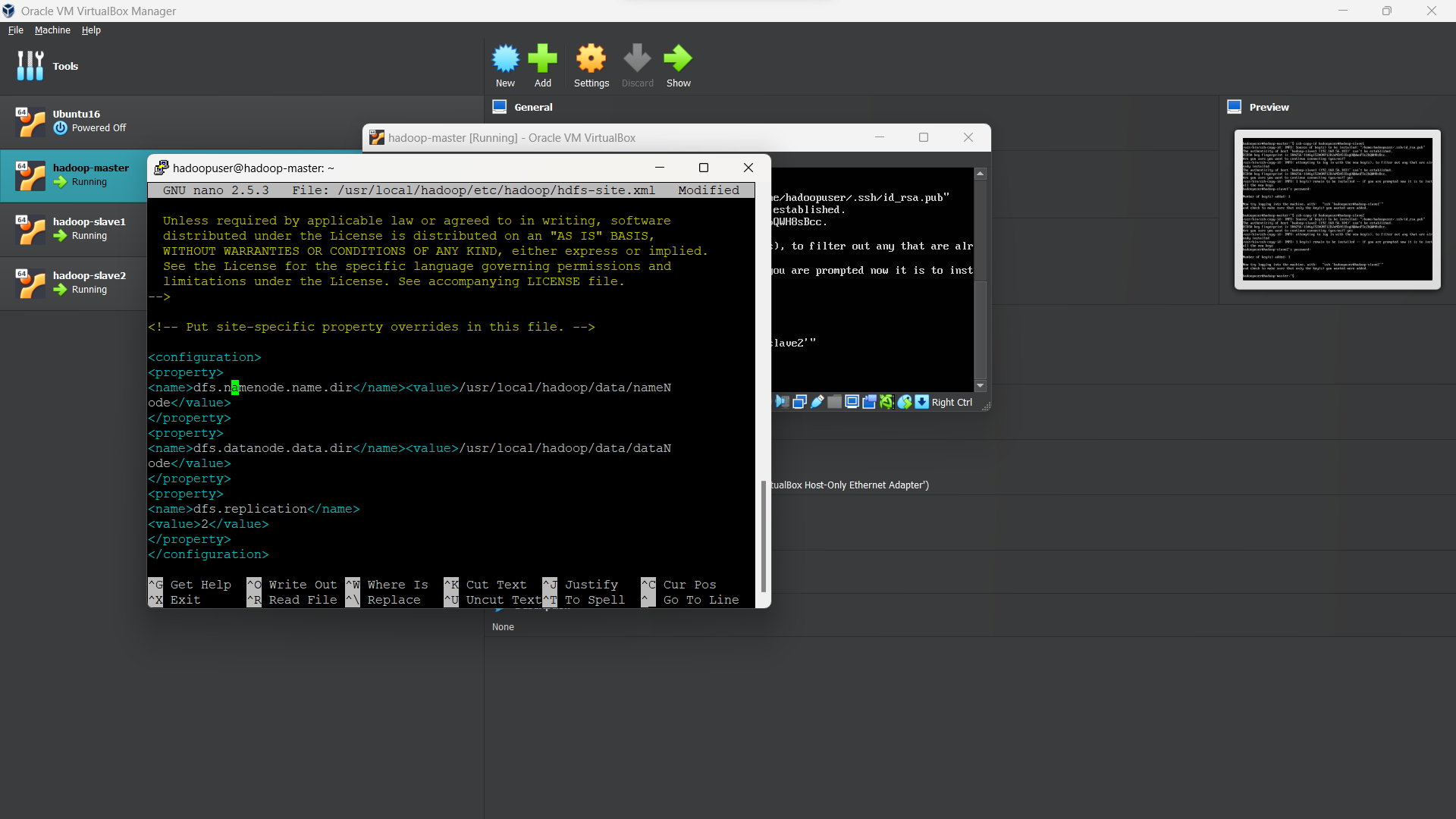
<name>dfs.replication</name>

<value>2</value>

</property>

</configuration>

Untuk keluar (menyimpan) dari editor Tekan “ctrl+X, Y, Enter”



Sama seperti kangkah sebelumnya bedanya adalah kalau disini adalah file hdfs-site.xml yang berfungsi untuk melakukan konfigurasi untuk peletakan ada di bagian paling akhir sesuai yang terterah pada gamabr di atas.

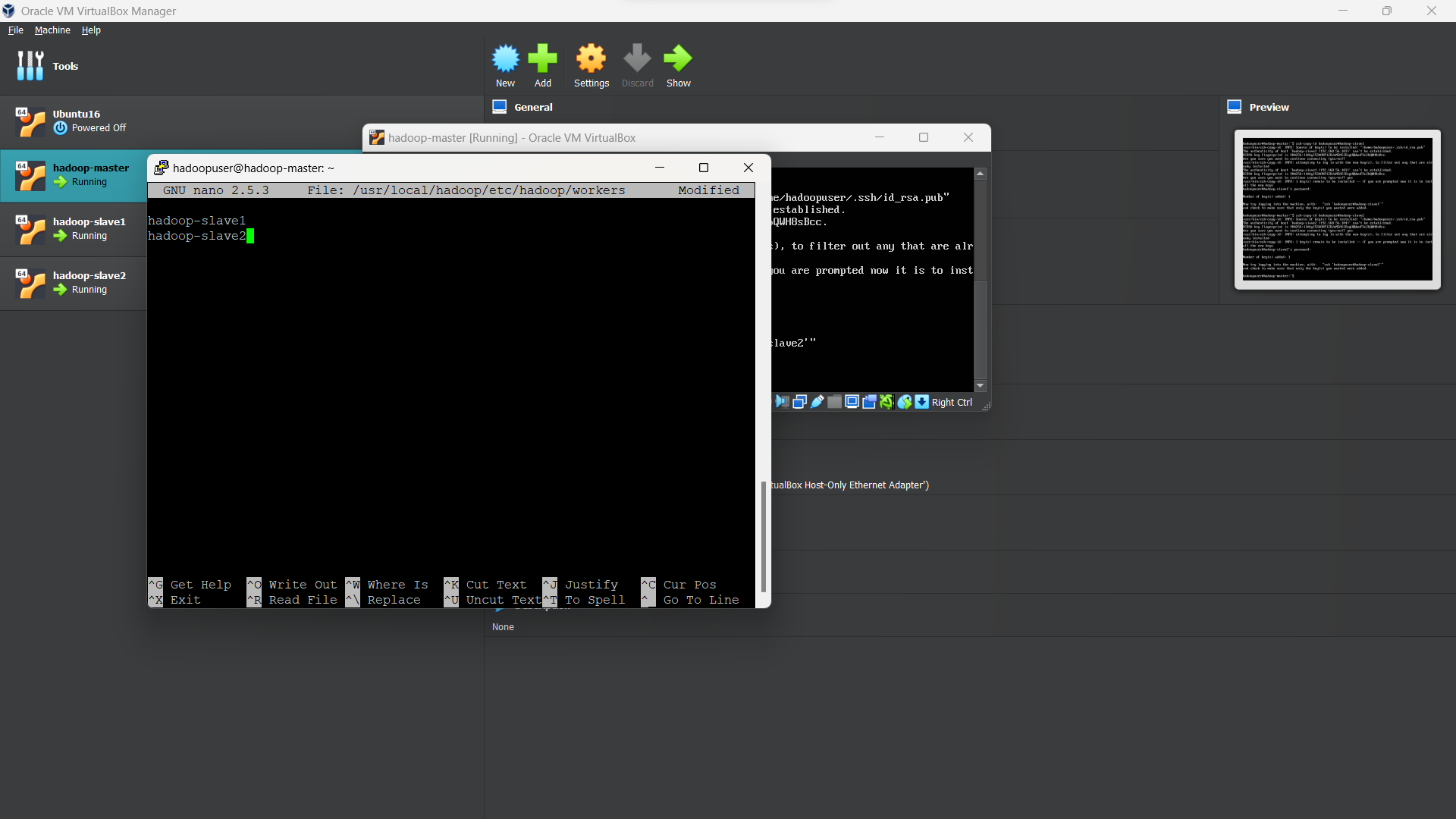
1. Pada **hadoop-master**, buka file **worker** untuk menambahkan node slave.

$ sudo nano /usr/local/hadoop/etc/hadoop/workers

hadoop-slave1

hadoop-slave2

Untuk keluar (menyimpan) dari editor Tekan “ctrl+X, Y, Enter”



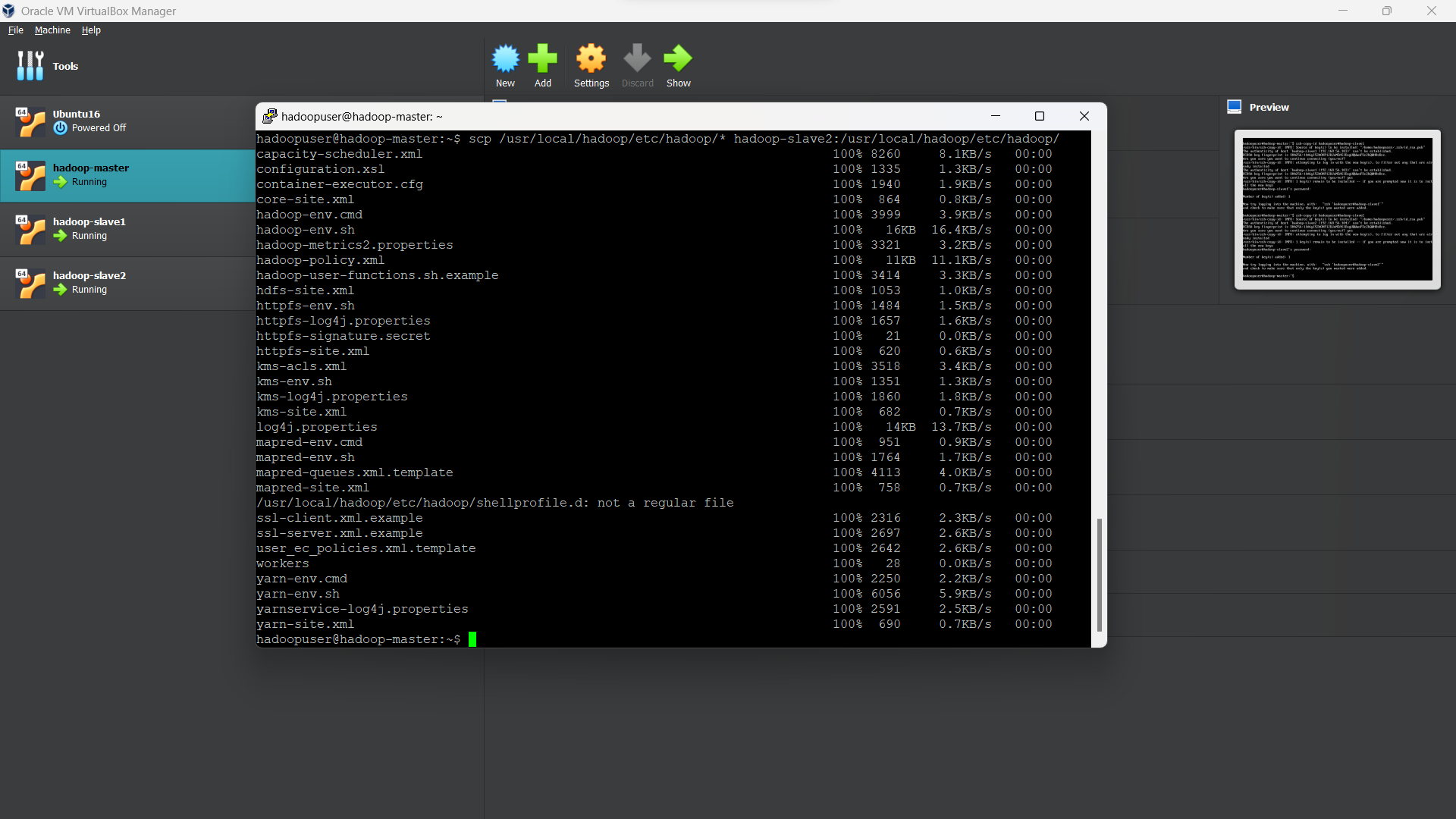
Disini kita akan membukan file worke untuk menambahkan node slave sesuian dengan nama vertual machine yang kita kloning dari hadoop-masternya.

1. Copy konfigurasi node **Hadoop-Master** ke node slaves, dengan menggunakan perintah berikut: (catatan: perintah ini dijalankan dalam satu baris, tidak ada enter)

$ scp /usr/local/hadoop/etc/hadoop/\* hadoopslave1:/usr/local/hadoop/etc/hadoop/

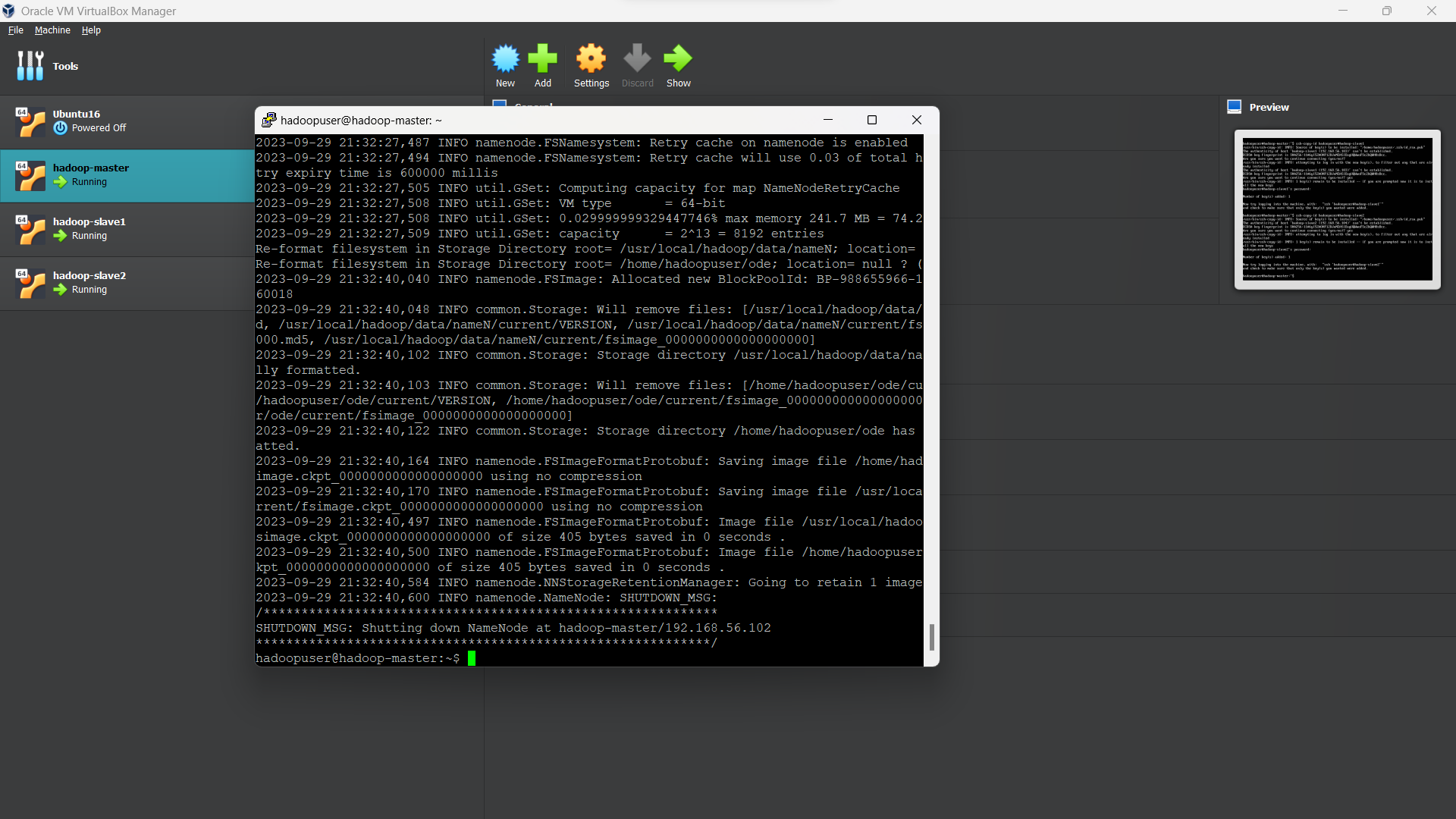
$ scp /usr/local/hadoop/etc/hadoop/\* hadoopslave2:/usr/local/hadoop/etc/hadoop/





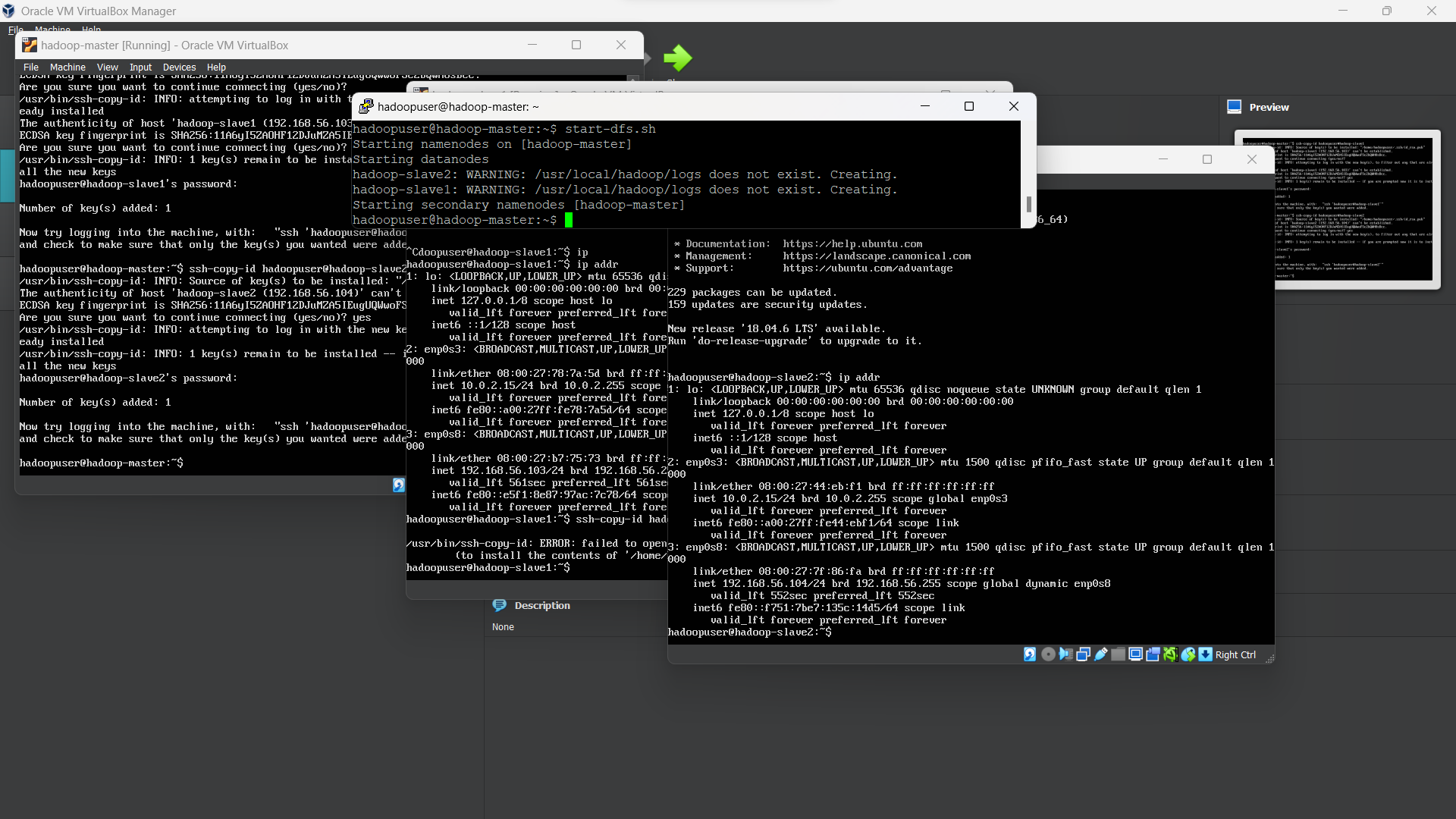
1. Selanjutnya pada **hadoop-master**, format HDFS file system, dengan menjalankan perintah berikut.

$ hdfs namenode -format



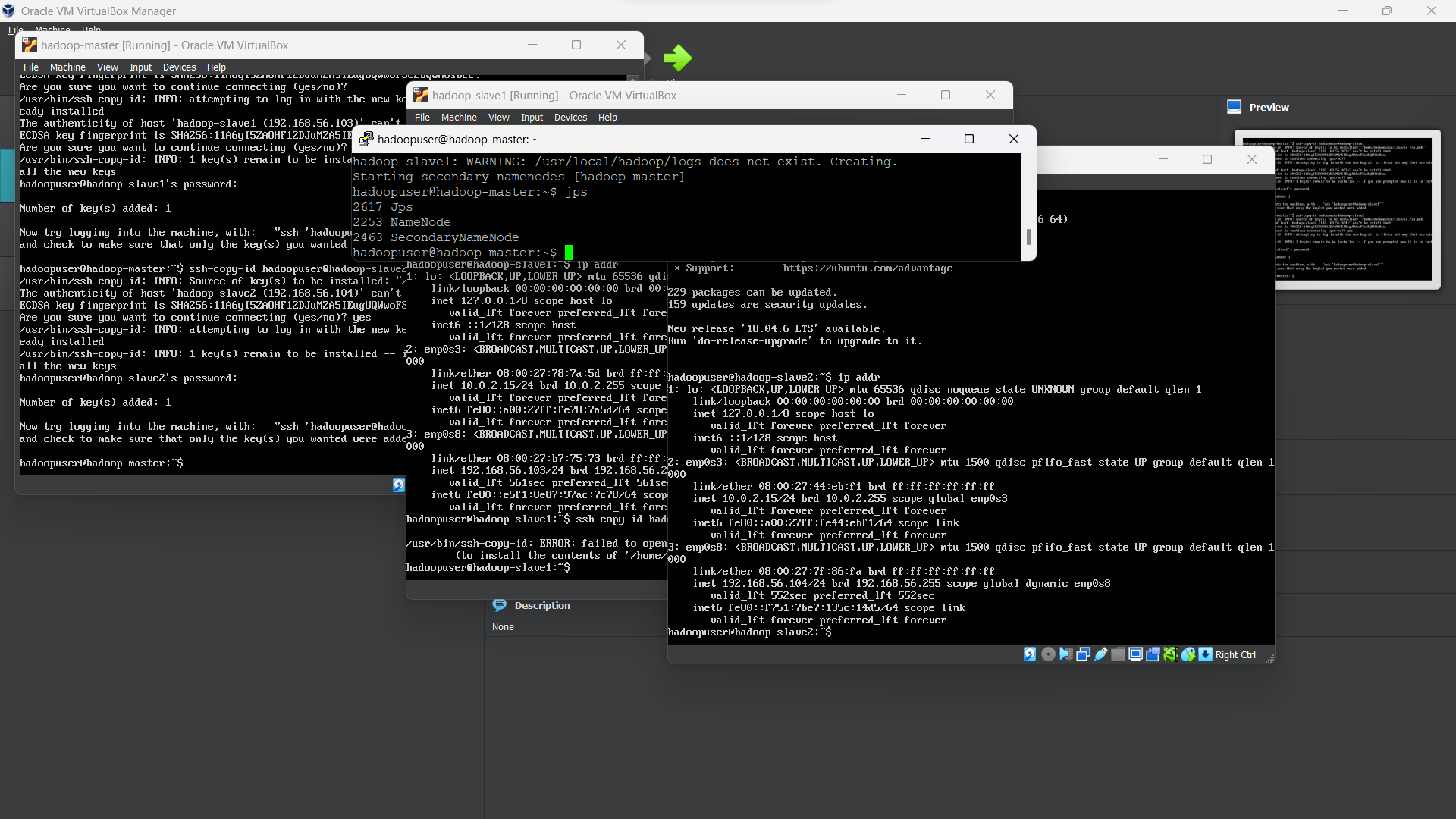
1. Pada **hadoo-master**, Start Hadoop cluster menggunakan scripts yang disediakan oleh hadoop. Untuk menjalankan script.

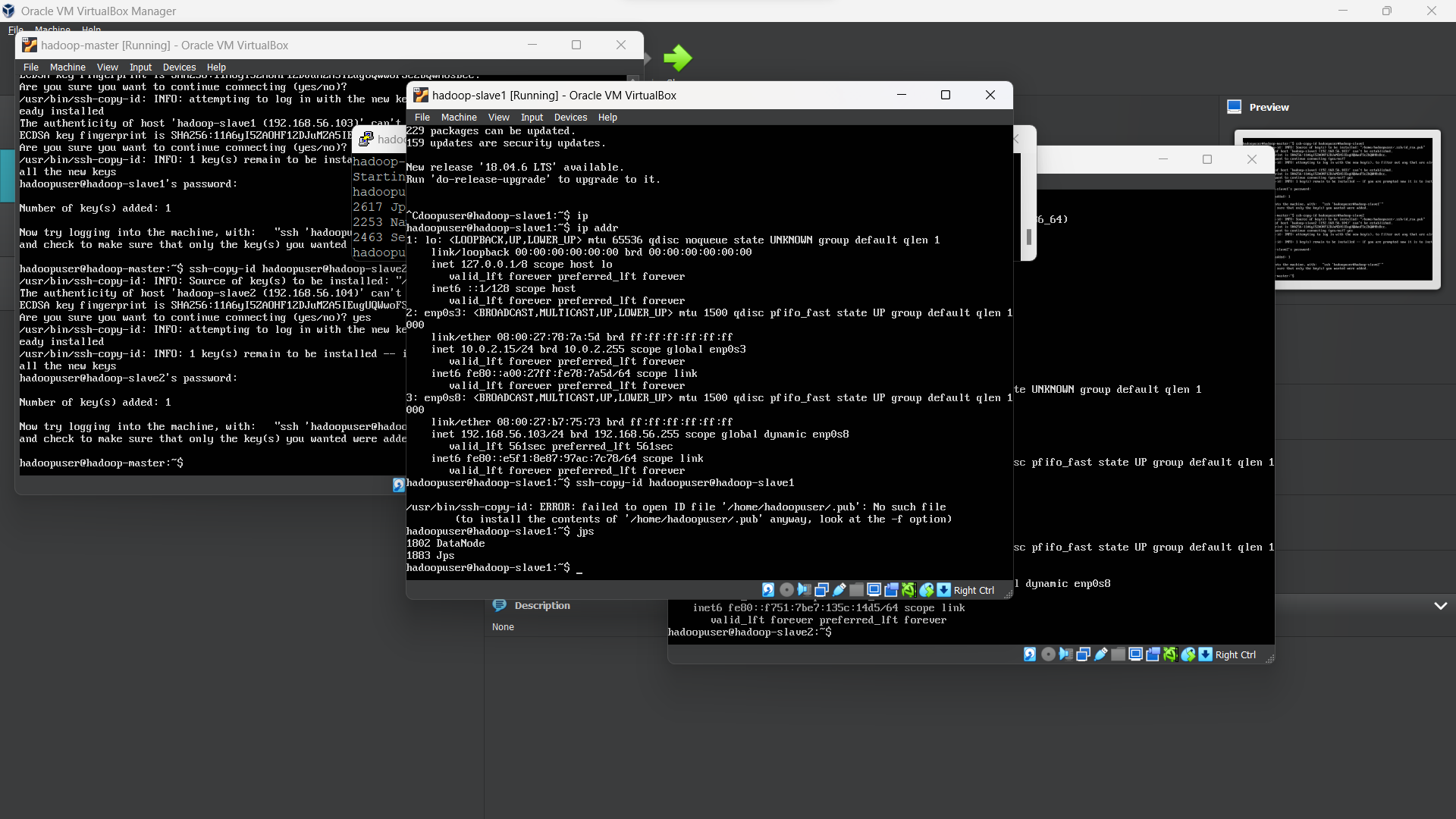
$ start-dfs.sh

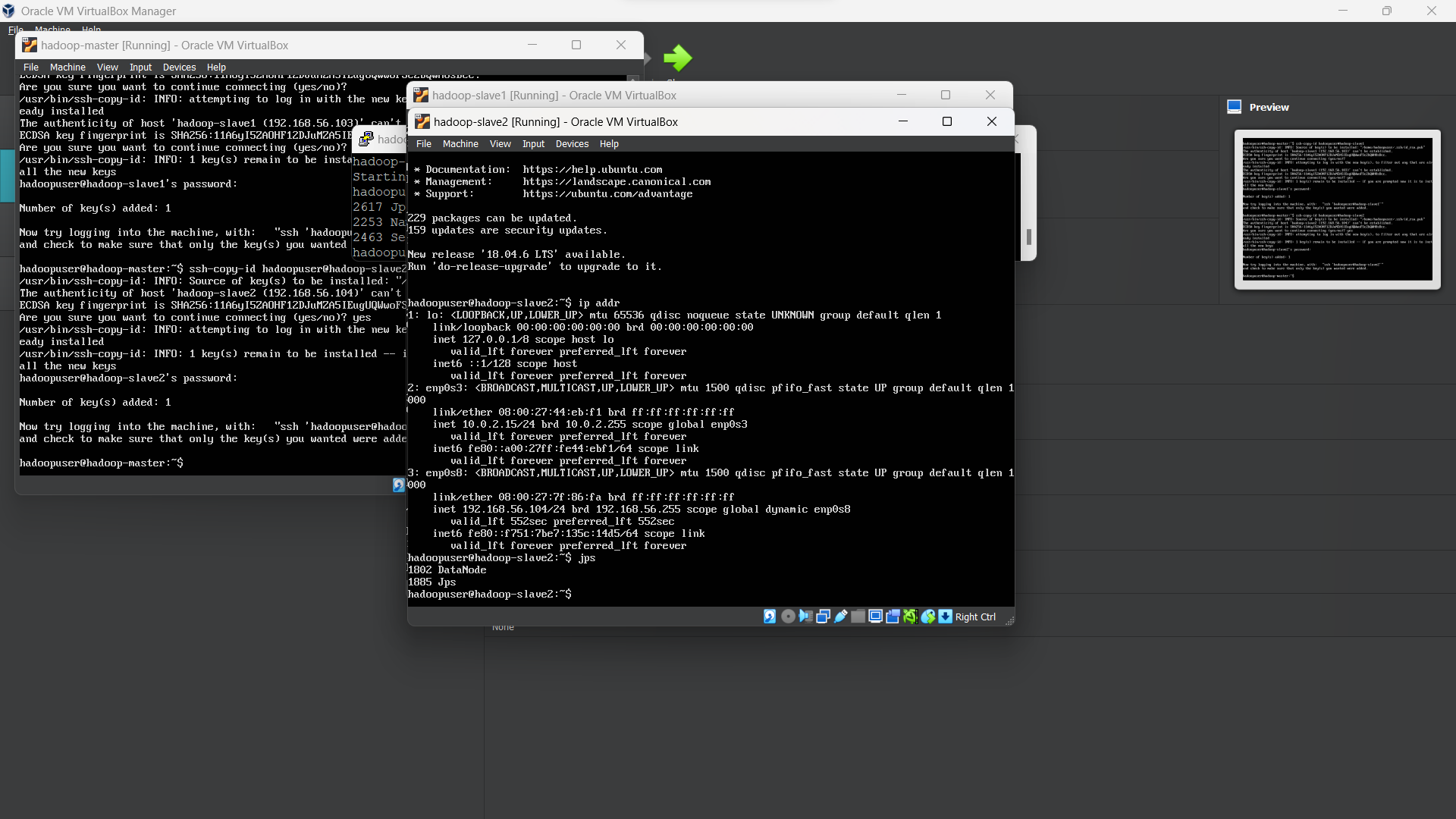


1. Untuk mengecek apakah hadoop sudah berjalan apa belum, gunakan perintah “**jps**” pada semua node Virtual Machine (hadoop-master, hadoop-slave1, & hadoop-slave2).

$ jps

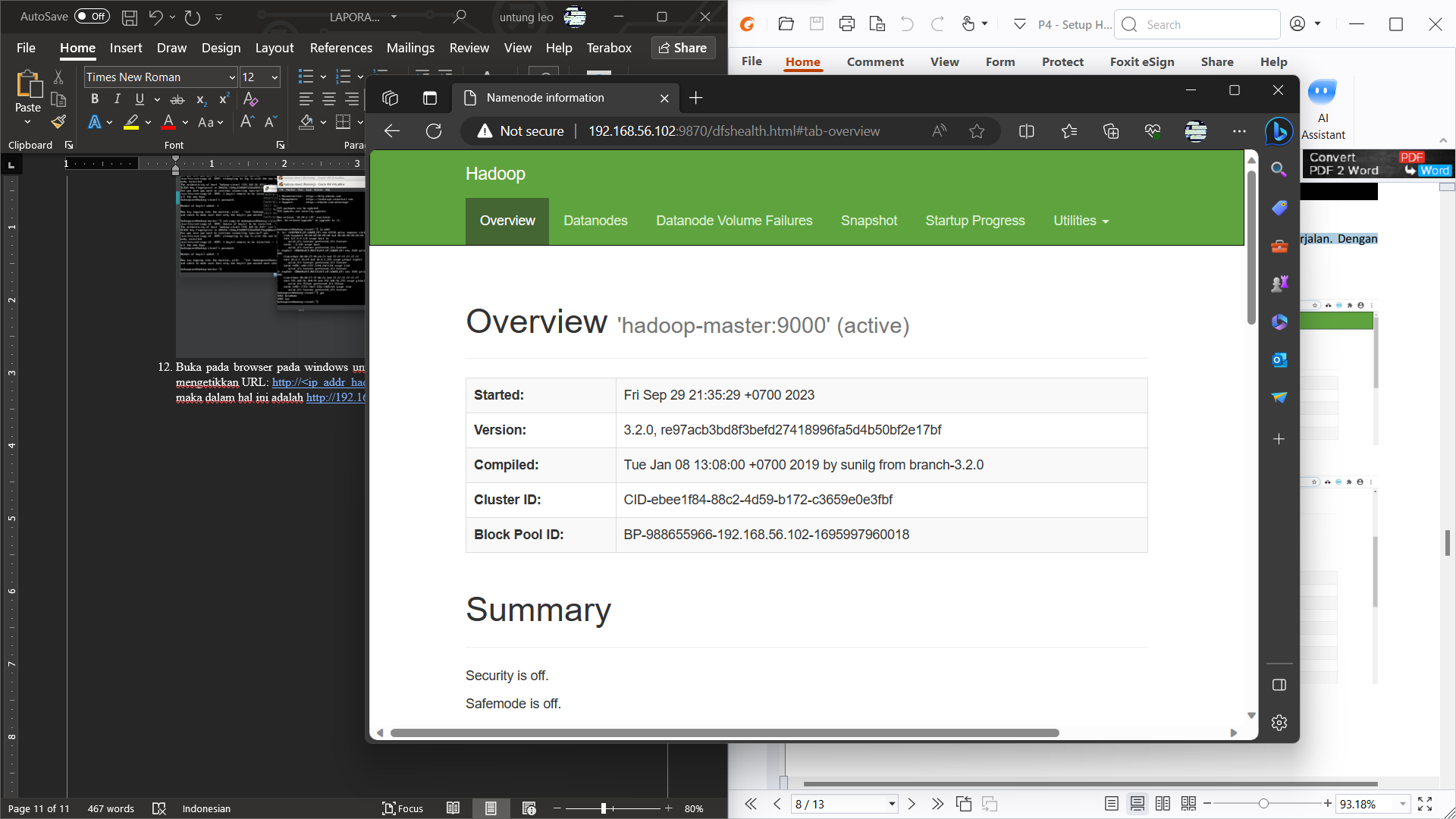




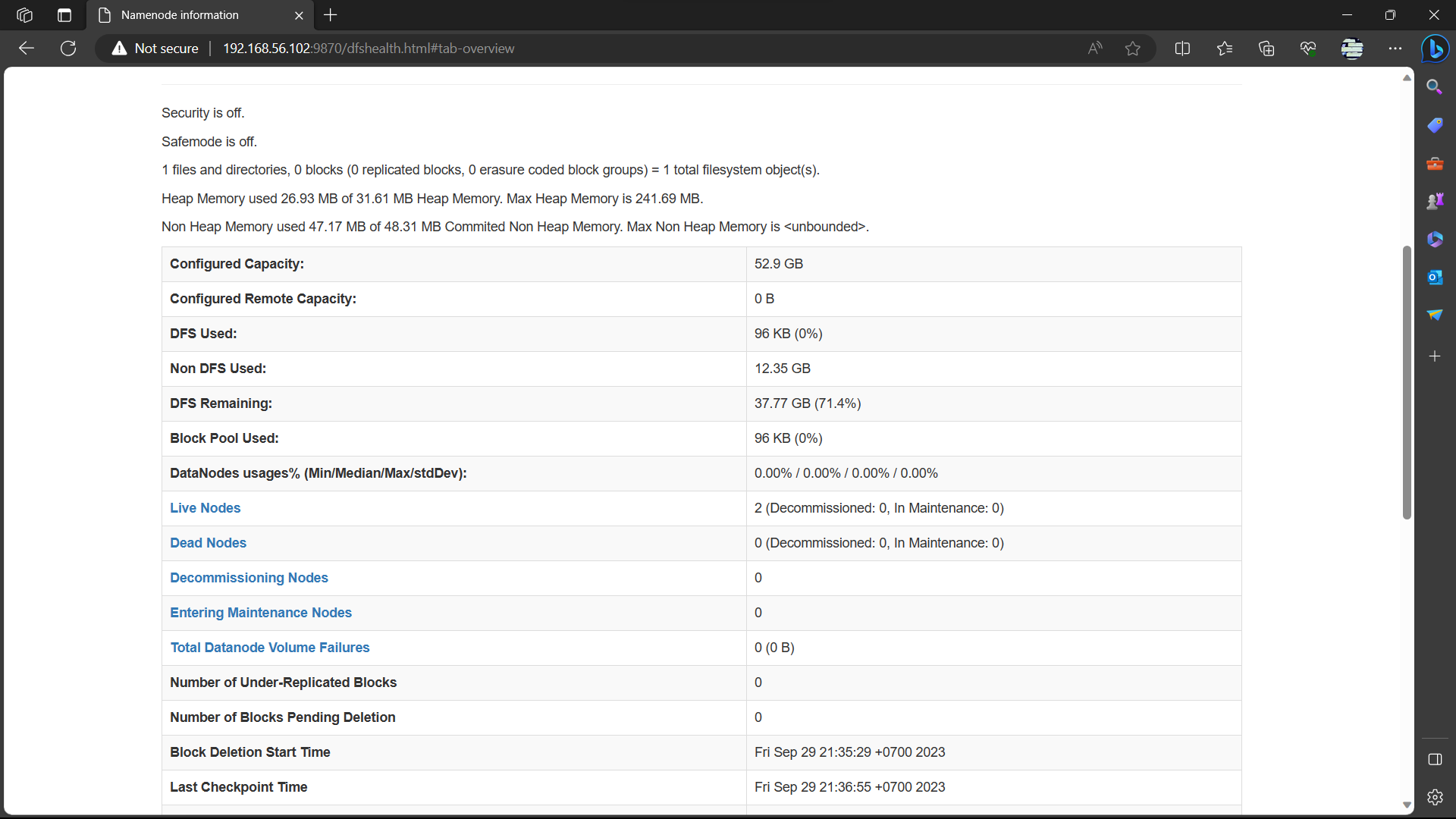


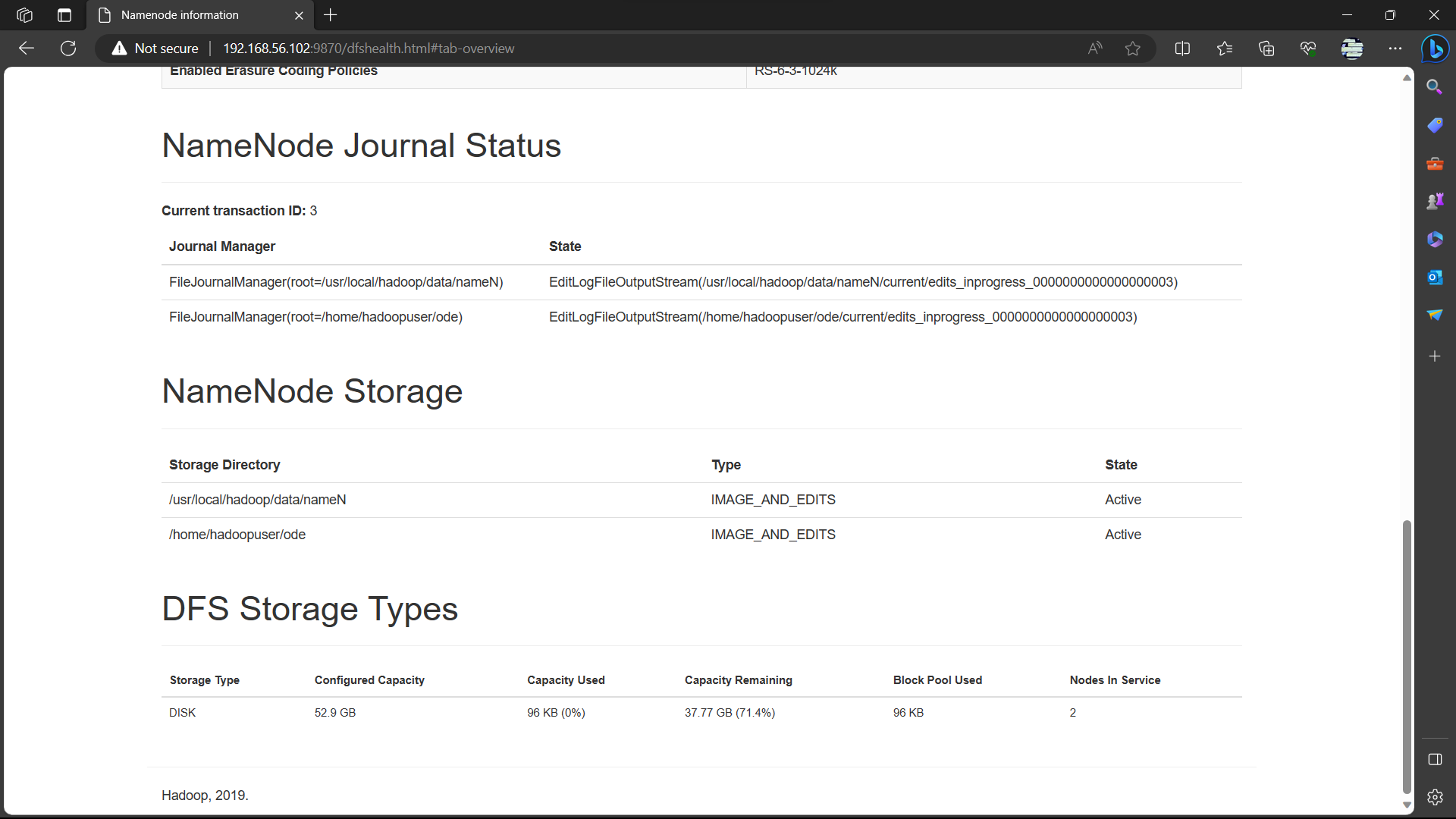
1. Buka pada browser pada windows untuk mengecek hadoop sudah berjalan. Dengan mengetikkan URL: http://<ip\_addr\_hadoop\_master>:9870/

maka dalam hal ini adalah <http://192.168.56.102:9870/>



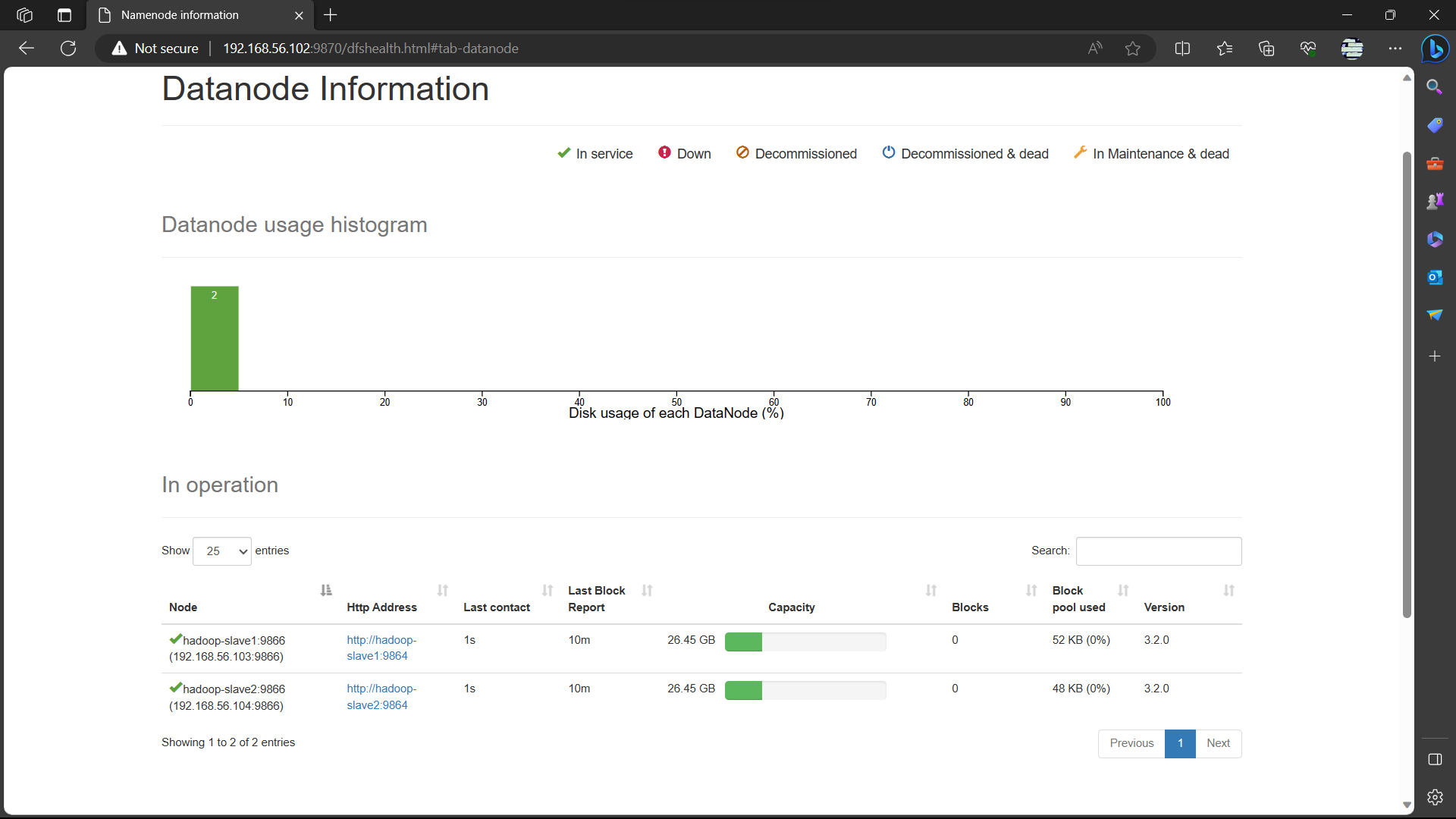
Scroll ke bawah kita akan mendapatkan informasi lebih banyak lagi





Klik pada menu “**datanode**”

Scroll ke bawah kita akan mendapatkan informasi lebih detail dari hadoop-slave.



1. Selanjutnya kita akan mengkonfigurasi **yarn**, pada **hadoop-master**. Modifikasi file “**.bashrc**” untuk membuat path direktori

$ nano .bashrc

export HADOOP\_HOME="/usr/local/hadoop"

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

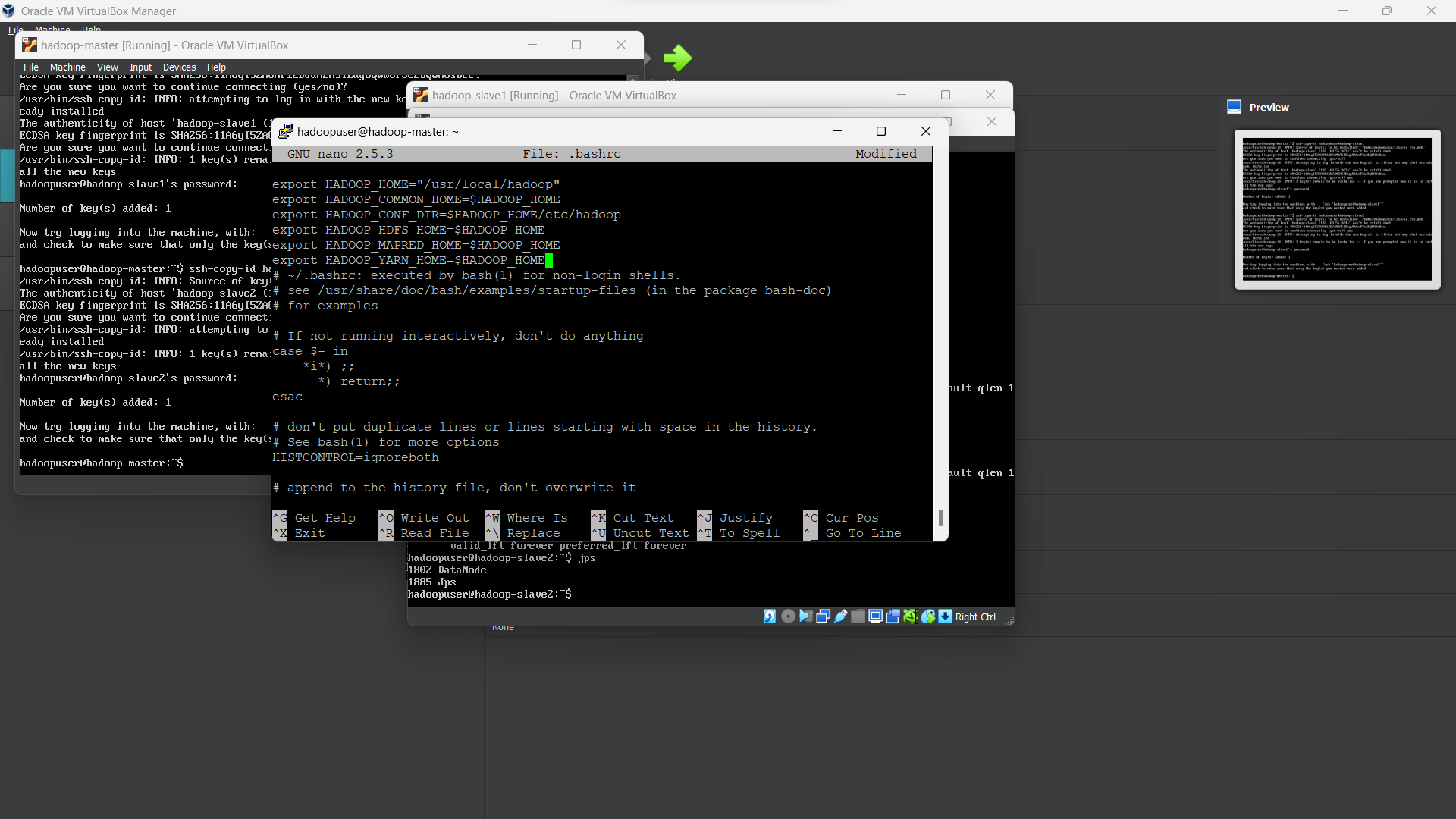
export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

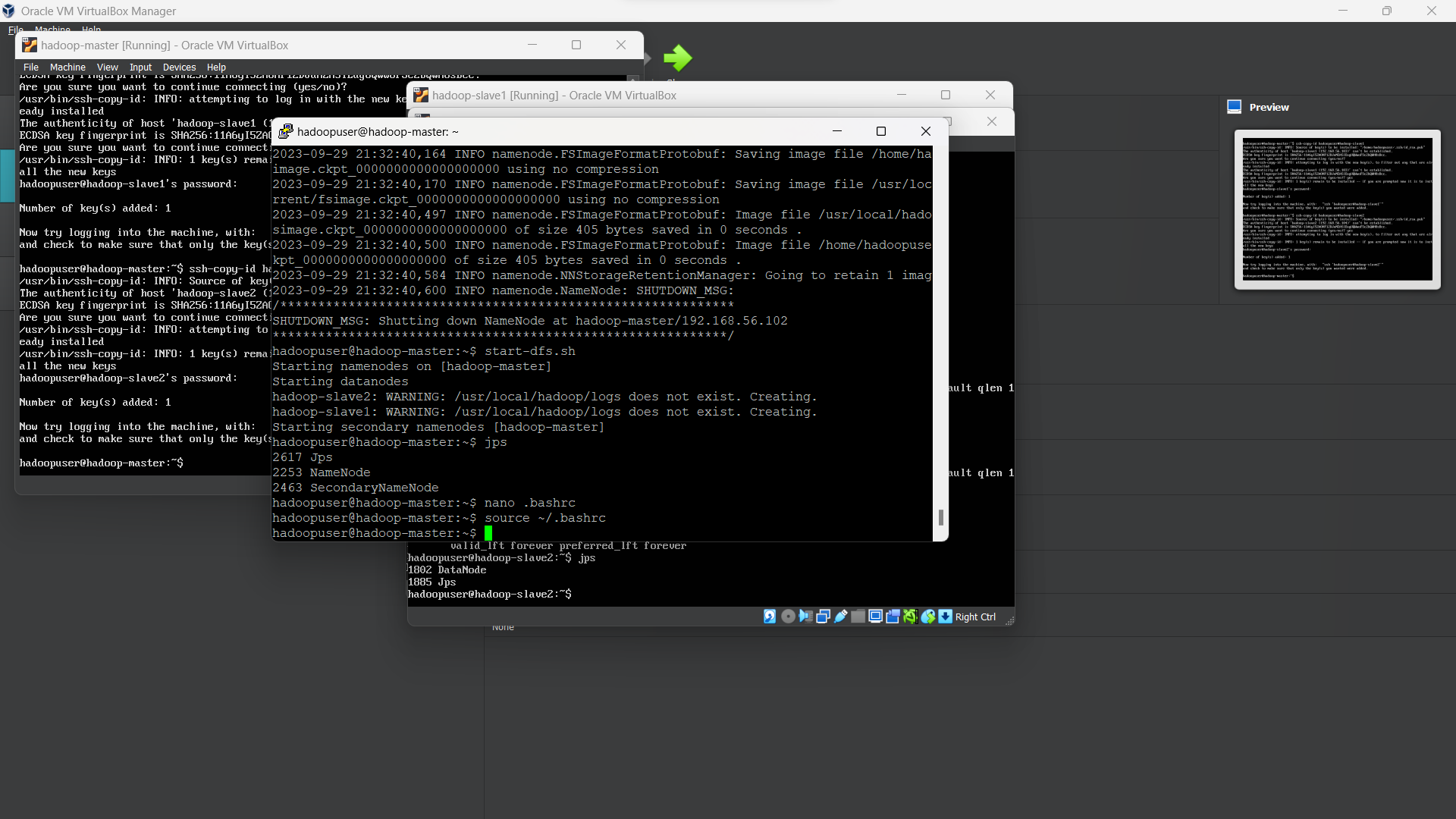
export HADOOP\_YARN\_HOME=$HADOOP\_HOME

Untuk keluar (menyimpan) dari editor Tekan “ctrl+X, Y, Enter”



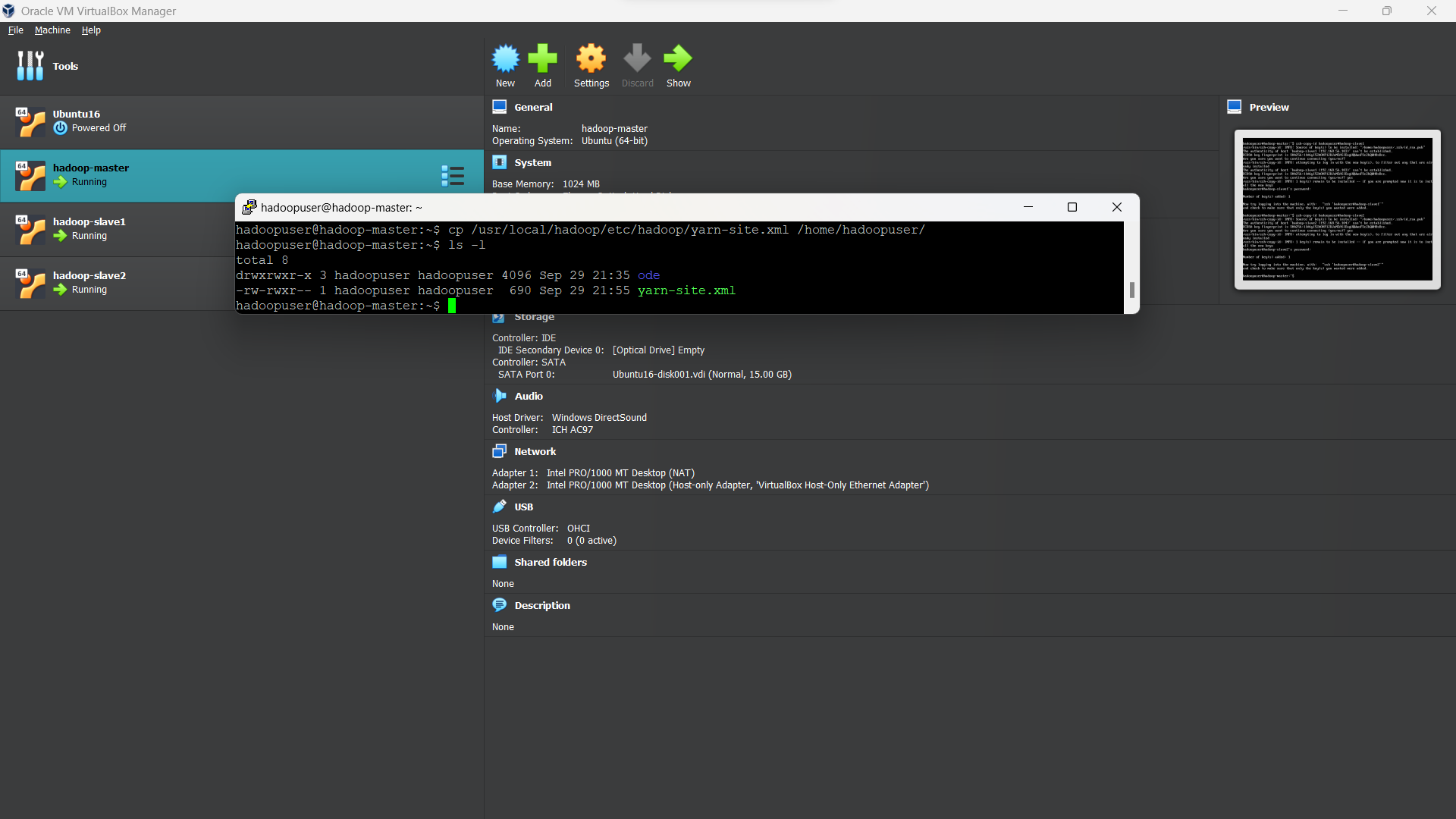
1. Refresh “**.bashrc**” dengan menggunakan perintah berikut.

$ source ~/.bashrc



1. Pada **hadoop-master** kita akan membuat file **yarn-site.xml** untuk kita gunakan mengkonfigurasi yarn-site.xml pada **hadoop-slave**. Langkah pertama adalah, kita akan mencopy file "**/usr/local/hadoop/etc/hadoop/yarn-site.xml**" ke direktori "**/home/hadoopuser**". Langkah kedua kita akan memodifikasi file "**/home/hadoopuser/yarn-site.xml**".

$ cp /usr/local/hadoop/etc/hadoop/yarn-site.xml /home/hadoopuser/



$ sudo nano /home/hadoopuser/yarn-site.xml

<configuration>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hadoop-master</value>

</property>

</configuration>

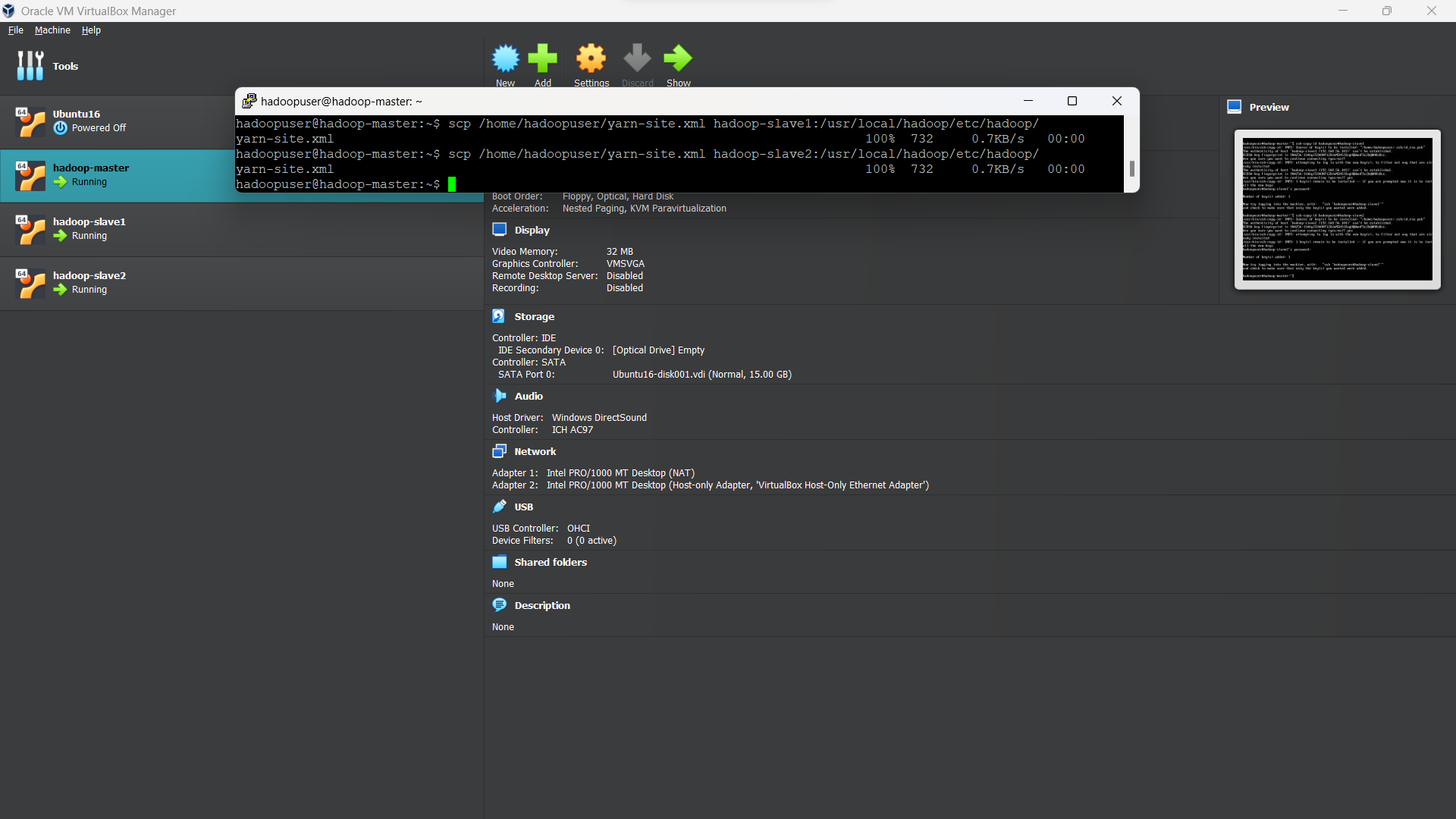
Untuk keluar (menyimpan) dari editor Tekan “ctrl+X, Y, Enter”



1. Copy file konfigurasi yarn-site.xml ke node slave. (catatan: perintah ini dijalankan dalam satu baris, tidak ada enter)

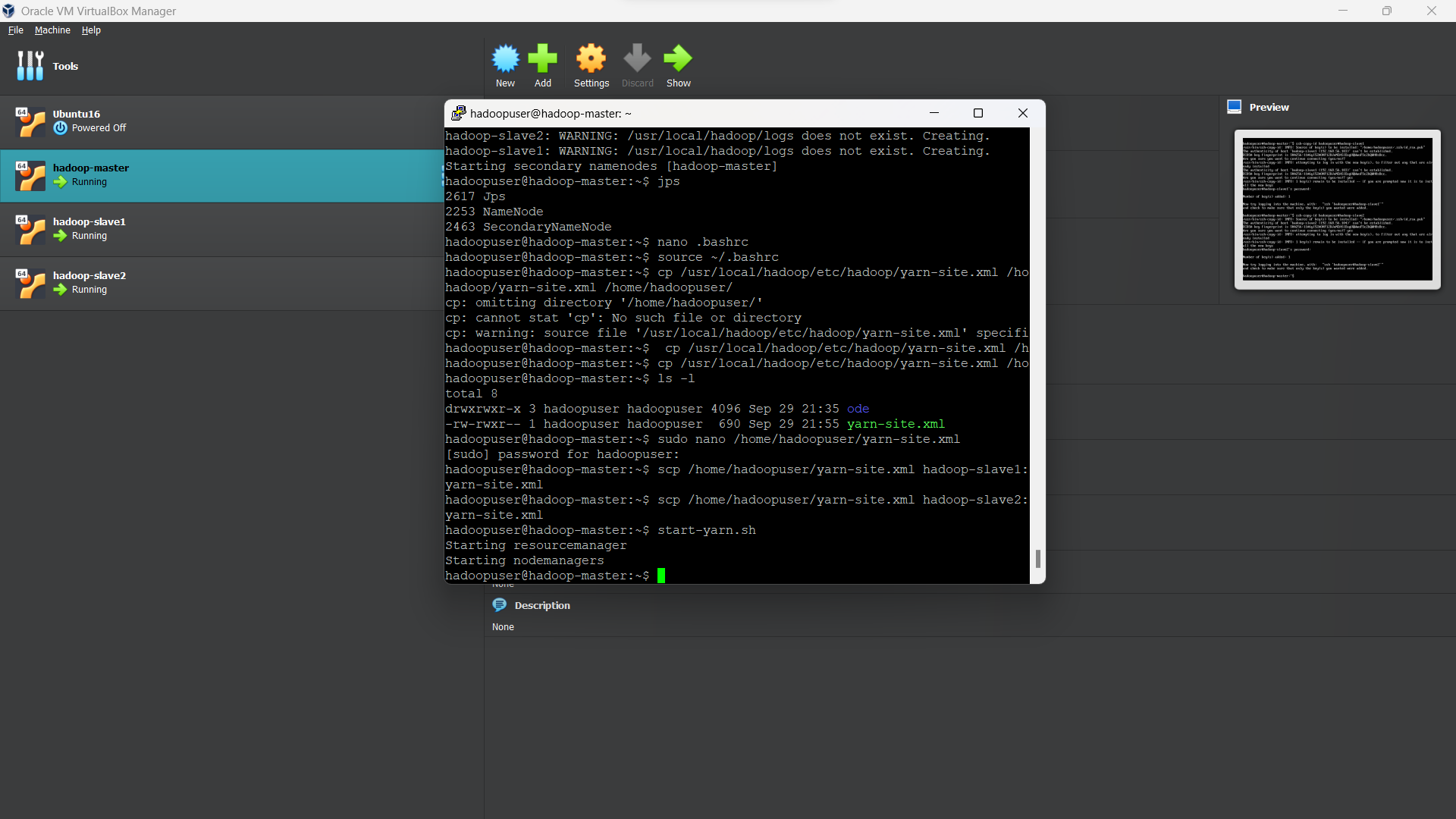
$ scp /home/hadoopuser/yarn-site.xml hadoop-slave1:/usr/local/hadoop/etc/hadoop/

$ scp /home/hadoopuser/yarn-site.xml hadoop-slave2:/usr/local/hadoop/etc/hadoop/



1. Pada **hadoop-master**, **Start yarn** menggunakan scripts yang disediakan oleh hadoop. Untuk menjalankan script.

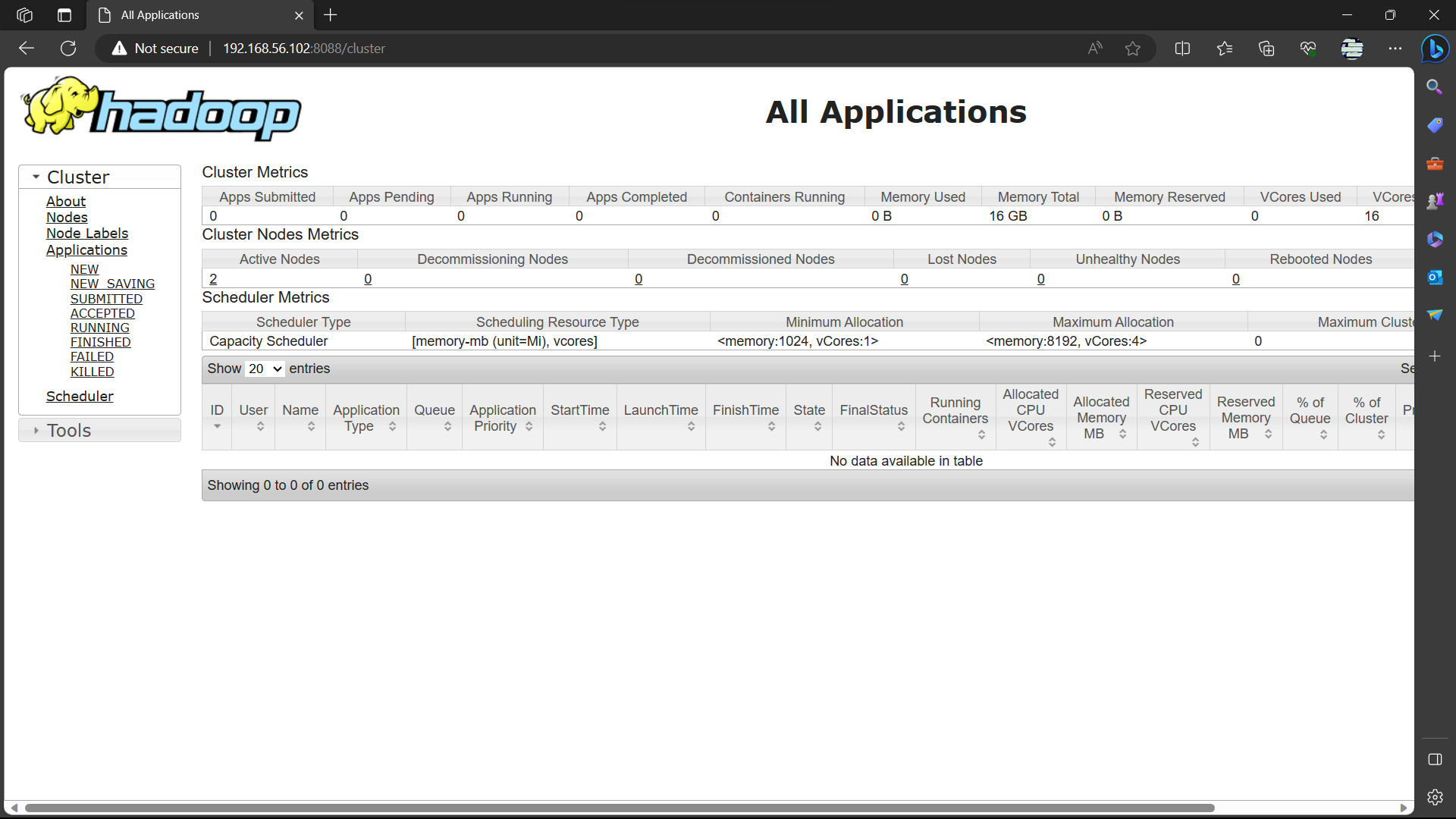
$ start-yarn.sh



1. Buka browser pada windows untuk mengecek apakah yarn sudah berjalan. Dengan mengetikkan URL: http://<ip\_addr\_hadoop\_master>:8088/

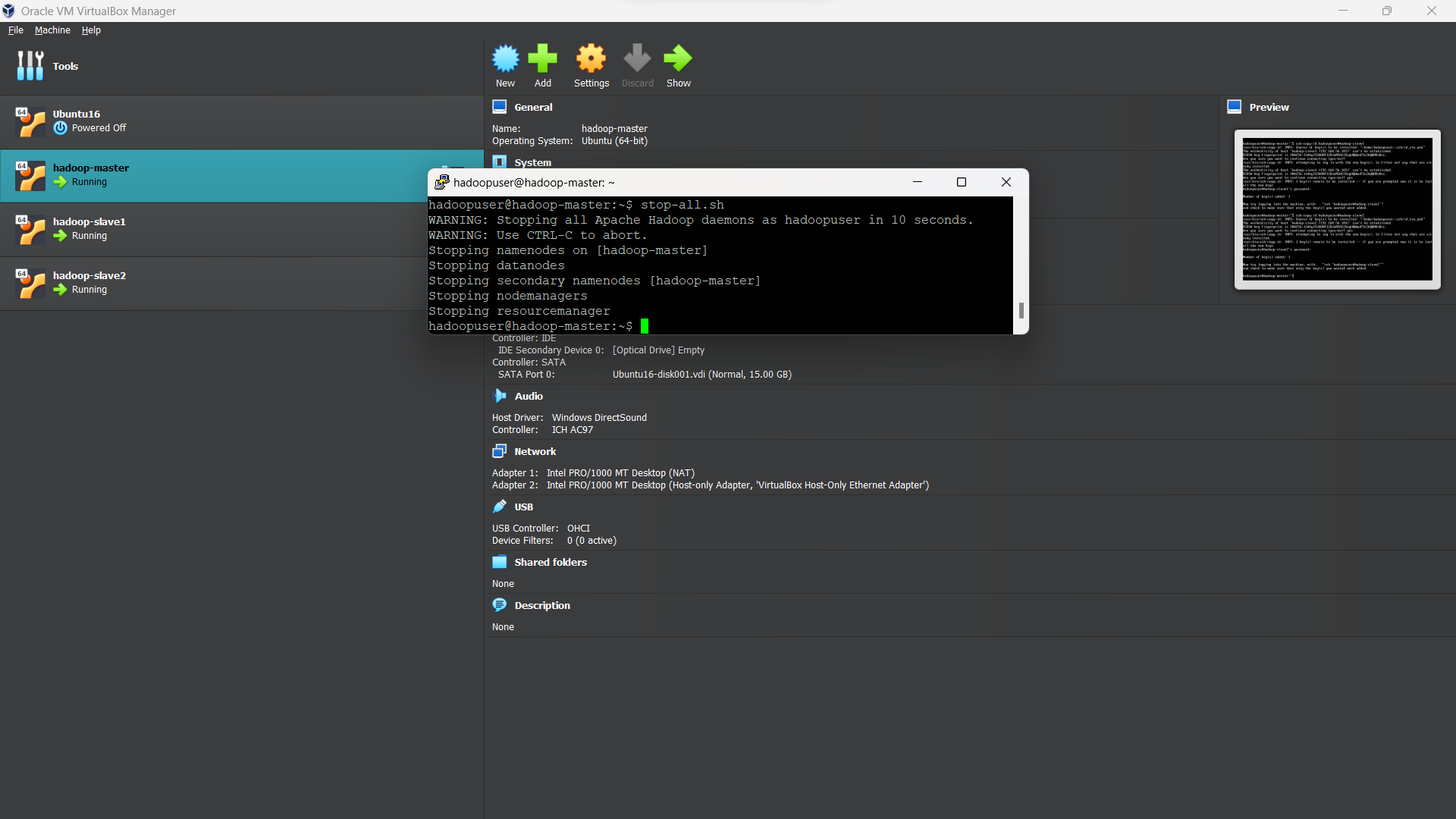
maka dalam hal ini adalah <http://192.168.56.102:8088/>

Dan kita akan dapat melihat ada **2 node** yang aktif, seperti diperlihatkan pada gambar



1. Untuk mematikan semua service hadoop, pada **hadoop-master** gunakan perintah

$ stop-all.sh



1. Matikan semua node dengan menggunakan perintah shutdown

$ sudo shutdown -h now

