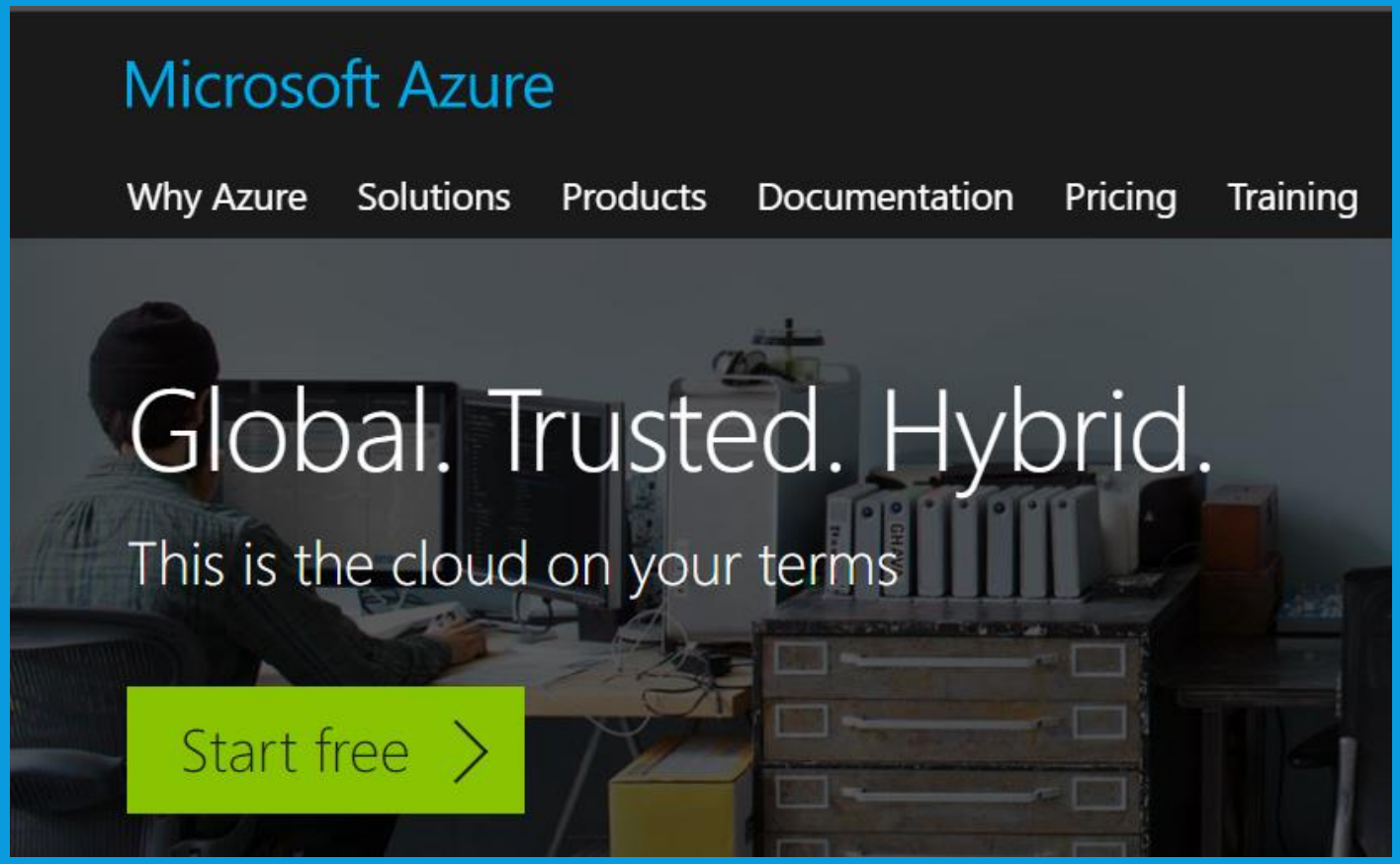
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

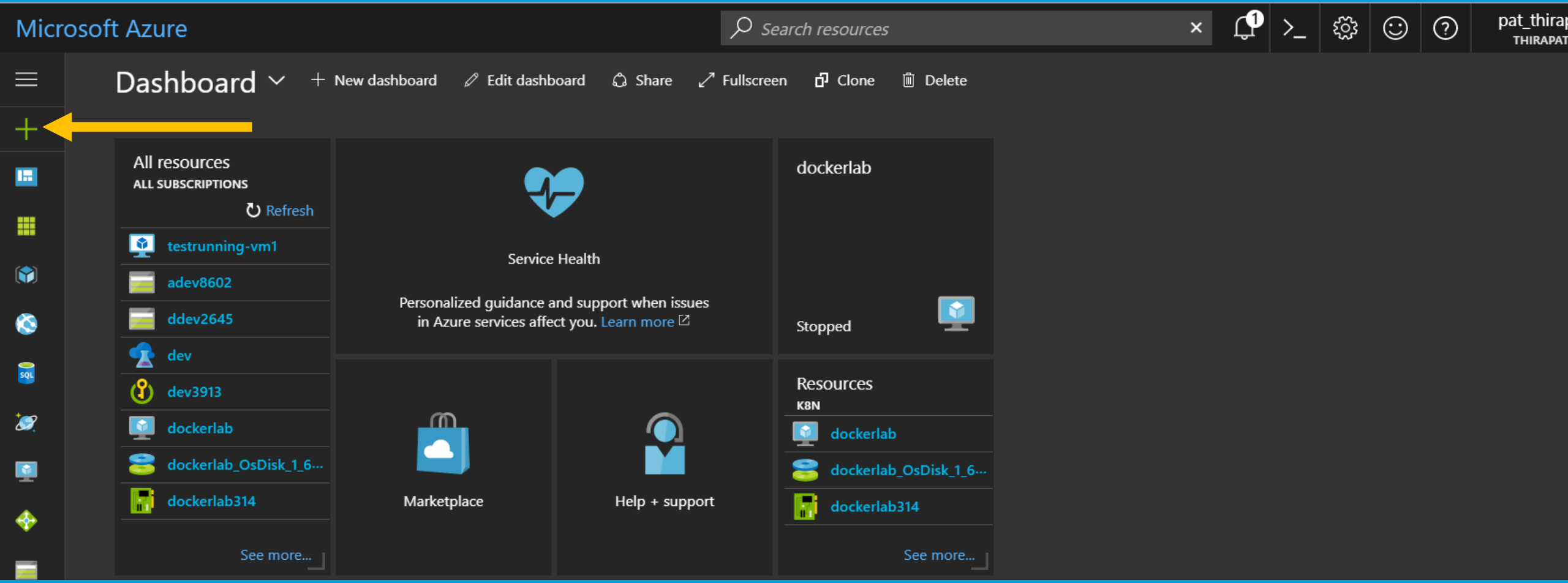
Thirapat Wiwittanaphorn

HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Step 1: Sign in to Azure Portal to get free Azure Credit**
- https://azure.microsoft.com/



HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

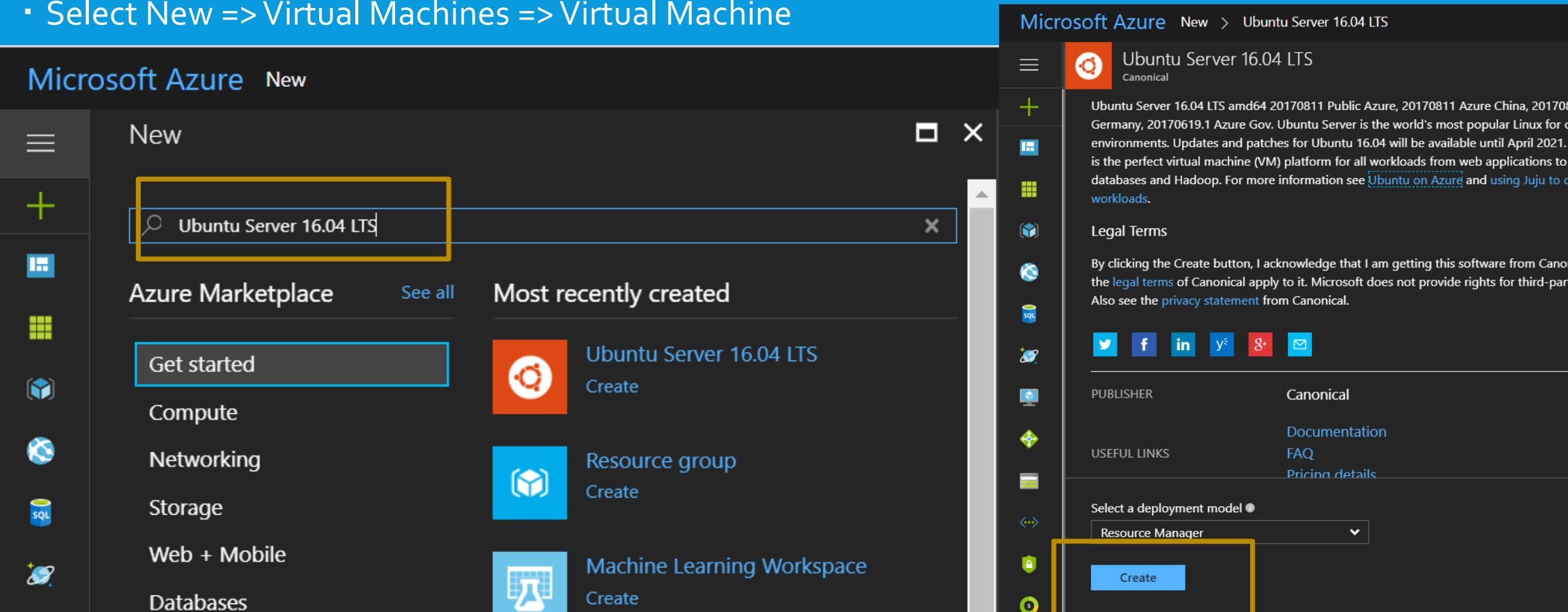


HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Launch Virtual machine Server**
- This lab use Ubuntu Server 16.04 LTS
  DS3_V2 Standard 4 Core, 14 GB memory,28 GB SSD

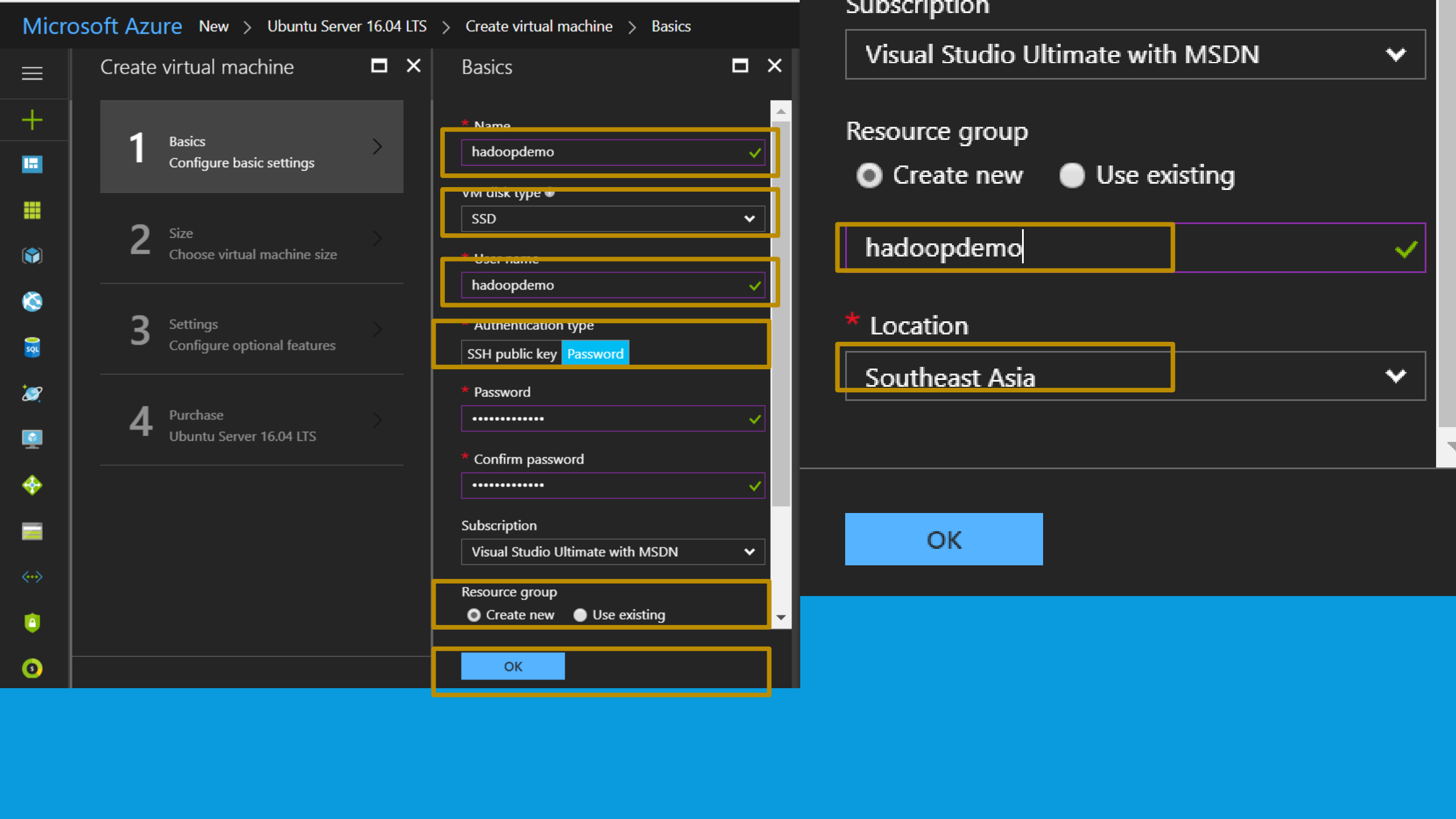
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

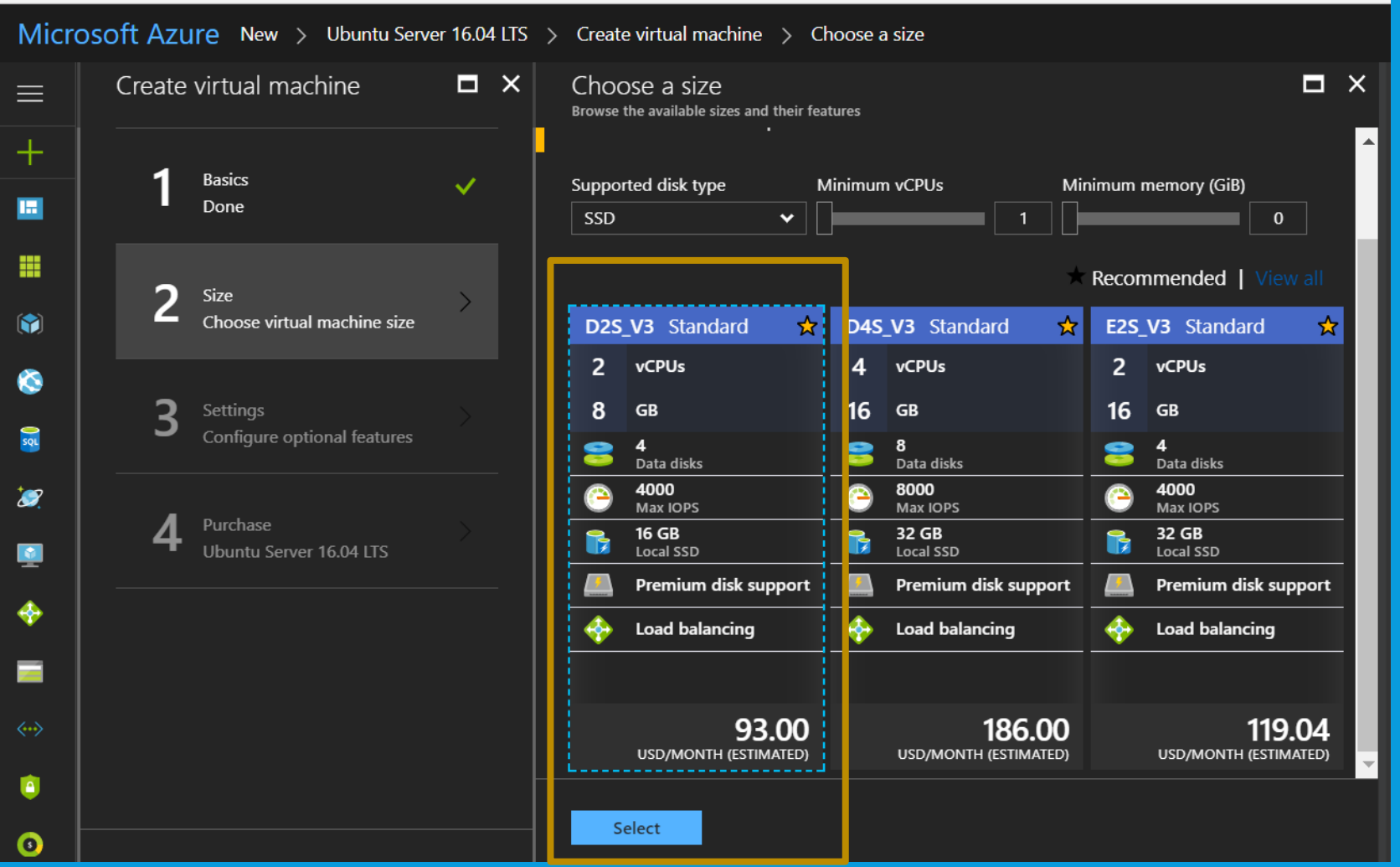
- Select New => Virtual Machines => Virtual Machine

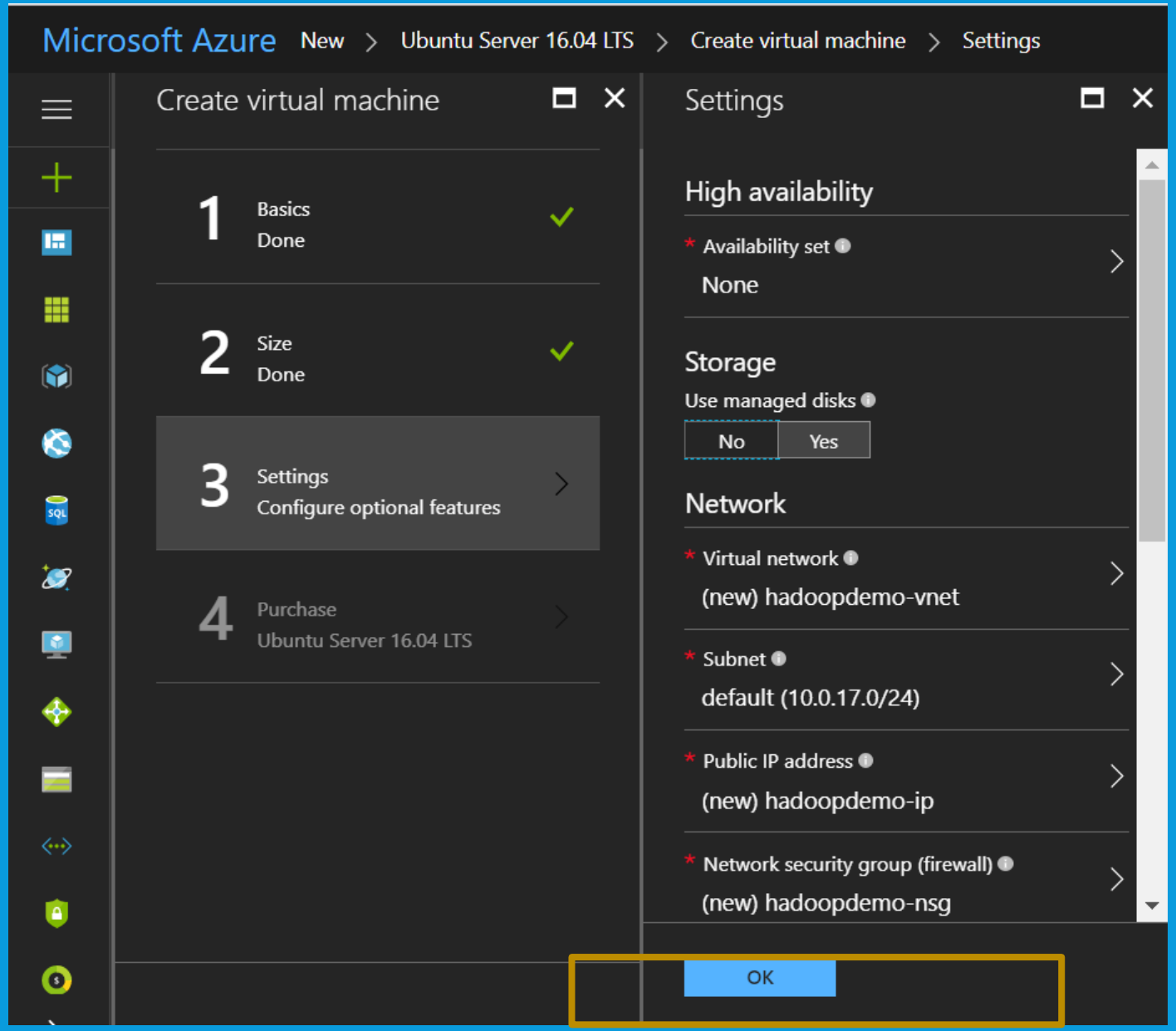


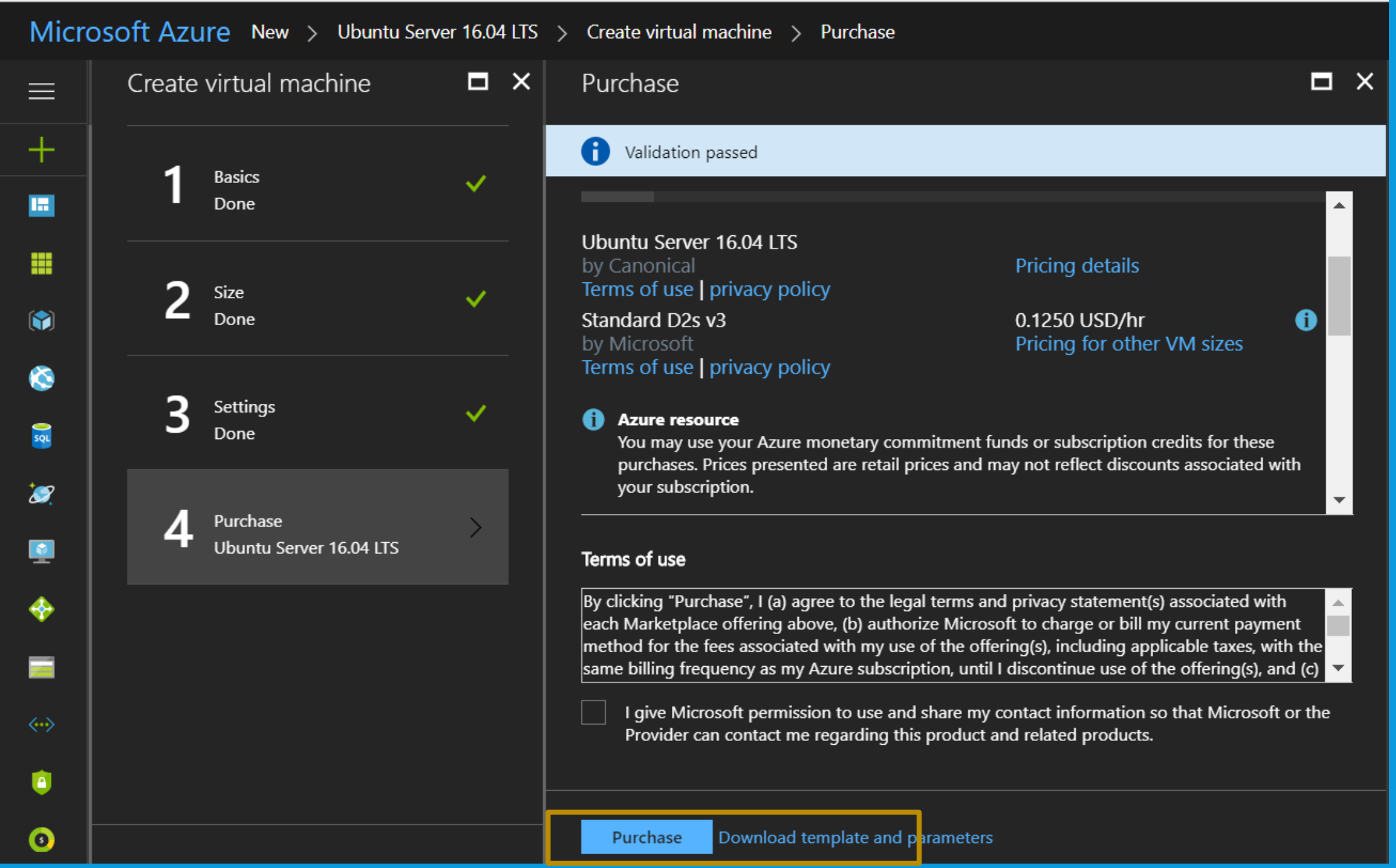
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Launch Virtual machine Server**
- On the Basics page, enter:
- a name for the VM
- a username for the Admin User
- the Authentication Type set to password
- a password
- a resource group name


Microsoft Azure
New
Ubuntu Server 16.04 LTS
Create virtual machine
Basics
Create virtual machine
1
Basics
Configure basic settings
2
Size
Choose virtual machine size
3
Settings
Configure optional features
4
Purchase
Ubuntu Server 16.04 LTS
Basics
hadoopdemo
SSD
hadoopdemo
SSH public key
Password
Password
Confirm password
Subscription
Visual Studio Ultimate with MSDN
Resource group
Create new
Use existing
OK
Subscription
Visual Studio Ultimate with MSDN
Resource group
Create new
Use existing
hadoopdemo
Location
Southeast Asia
OK

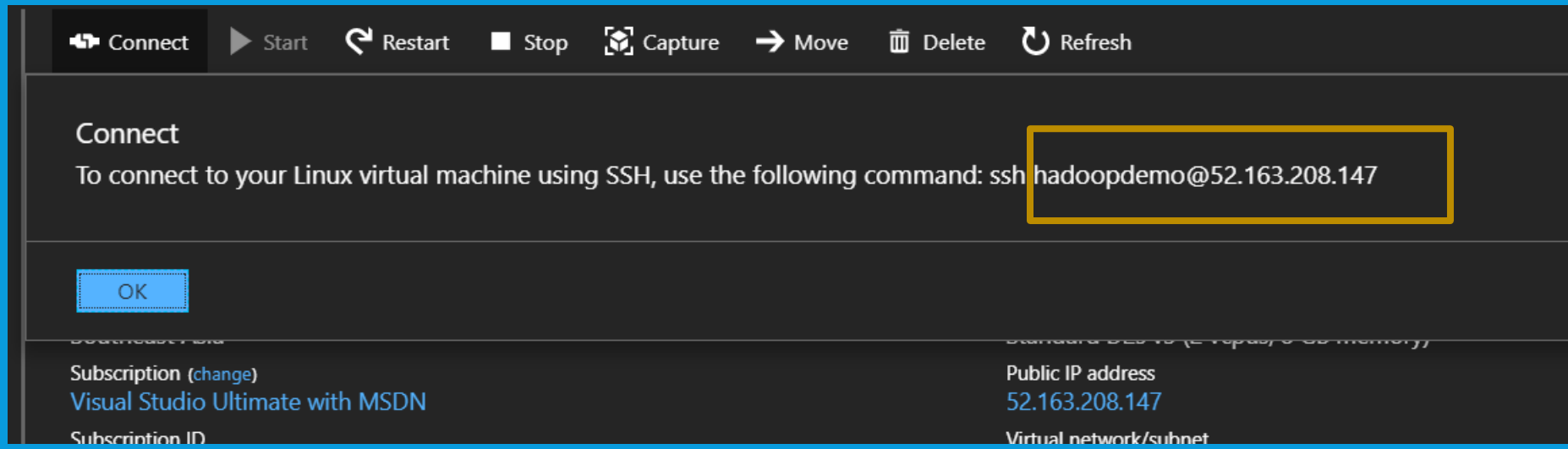

Microsoft Azure
New
Ubuntu Server 16.04 LTS
Create virtual machine
Choose a size
Create virtual machine
1
Basics
Done
2
Size
Choose virtual machine size
3
Settings
Configure optional features
4
Purchase
Ubuntu Server 16.04 LTS
Choose a size
Browse the available sizes and their features
Supported disk type
SSD
Minimum vCPUs
1
Minimum memory (GiB)
0
Recommended | View all
D2S_V3 Standard
2 vCPUs
8 GB
4 Data disks
4000 Max IOPS
16 GB Local SSD
Premium disk support
Load balancing
93.00
USD/MONTH (ESTIMATED)
D4S_V3 Standard
4 vCPUs
16 GB
8 Data disks
8000 Max IOPS
32 GB Local SSD
Premium disk support
Load balancing
186.00
USD/MONTH (ESTIMATED)
E2S_V3 Standard
2 vCPUs
16 GB
4 Data disks
4000 Max IOPS
32 GB Local SSD
Premium disk support
Load balancing
119.04
USD/MONTH (ESTIMATED)
Select


Microsoft Azure
New
Ubuntu Server 16.04 LTS
Create virtual machine
Settings
Create virtual machine
1
Basics
Done
2
Size
Done
3
Settings
Configure optional features
4
Purchase
Ubuntu Server 16.04 LTS
Settings
High availability
* Availability set
None
Storage
Use managed disks
No
Yes
Network
* Virtual network
(new) hadoopdemo-vnet
* Subnet
default (10.0.17.0/24)
* Public IP address
(new) hadoopdemo-ip
* Network security group (firewall)
(new) hadoopdemo-nsg
OK


Microsoft Azure
New
Ubuntu Server 16.04 LTS
Create virtual machine
Purchase
Create virtual machine
1
Basics
Done
2
Size
Done
3
Settings
Done
4
Purchase
Ubuntu Server 16.04 LTS
Purchase
Validation passed
Ubuntu Server 16.04 LTS
by Canonical
Terms of use | privacy policy
Pricing details
Standard D2s v3
by Microsoft
Terms of use | privacy policy
0.1250 USD/hr
Pricing for other VM sizes
Azure resource
You may use your Azure monetary commitment funds or subscription credits for these purchases. Prices presented are retail prices and may not reflect discounts associated with your subscription.
Terms of use
By clicking "Purchase", I (a) agree to the legal terms and privacy statement(s) associated with each Marketplace offering above, (b) authorize Microsoft to charge or bill my current payment method for the fees associated with my use of the offering(s), including applicable taxes, with the same billing frequency as my Azure subscription, until I discontinue use of the offering(s), and (c)
I give Microsoft permission to use and share my contact information so that Microsoft or the Provider can contact me regarding this product and related products.
Purchase
Download template and parameters

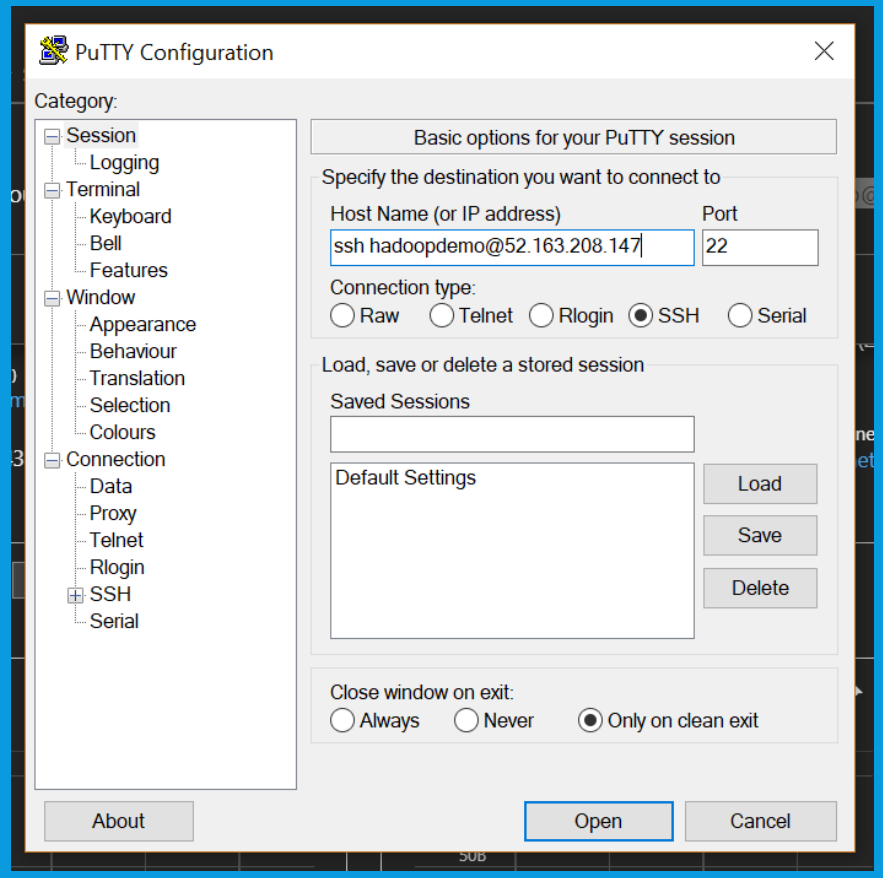
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- Get the IP address



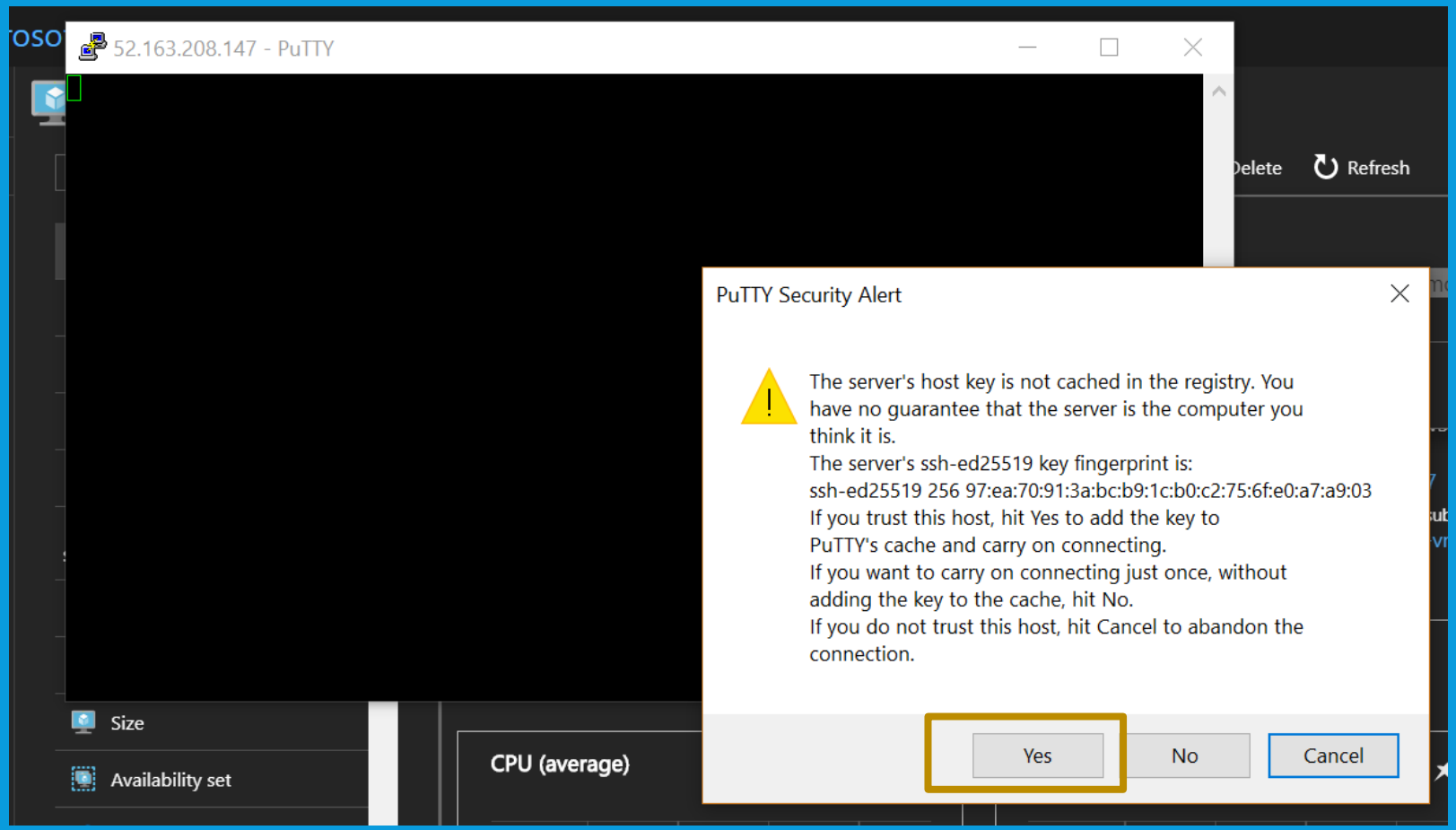
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Connect to an instance from Windows using Putty**



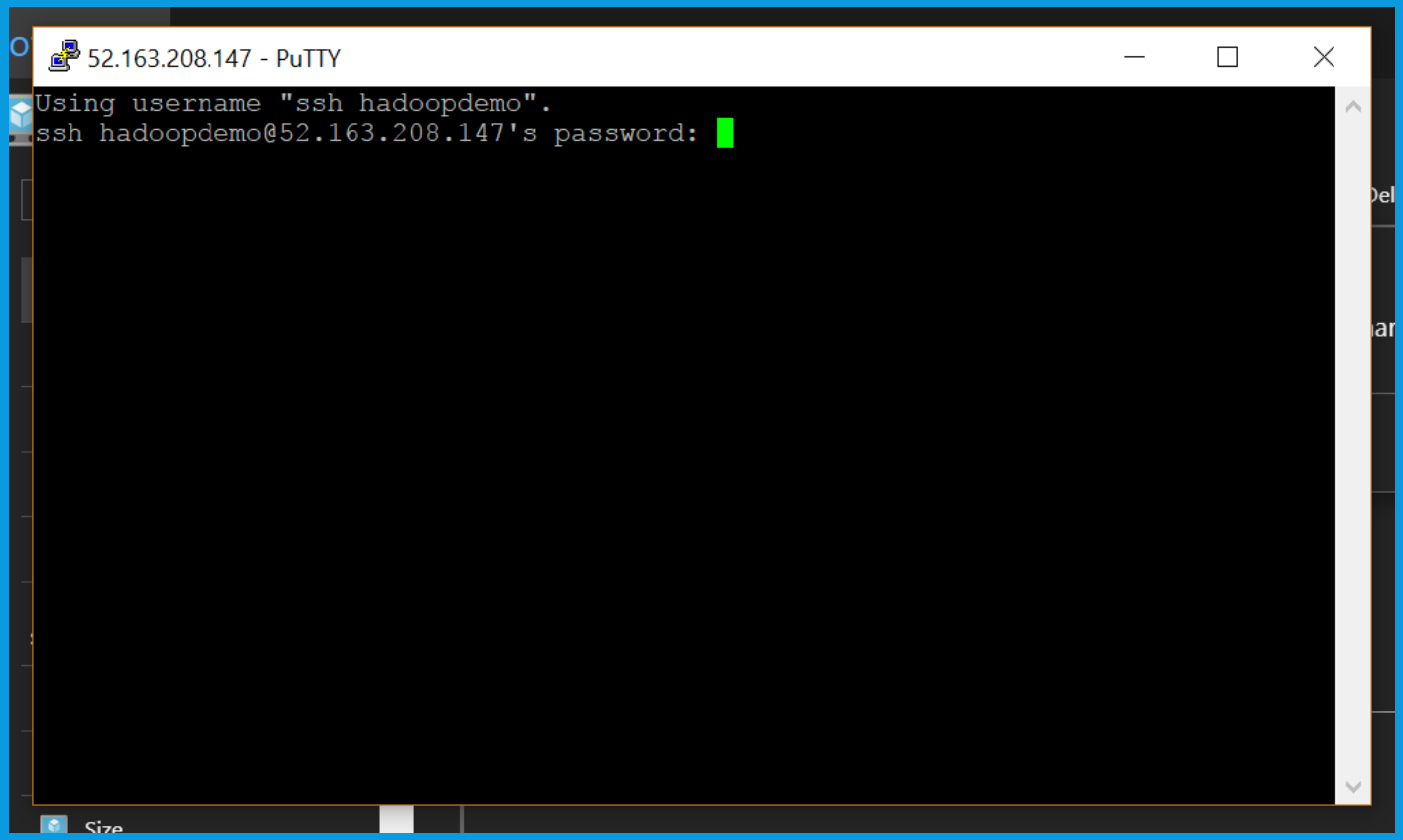
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Connect to an instance from Windows using Putty**



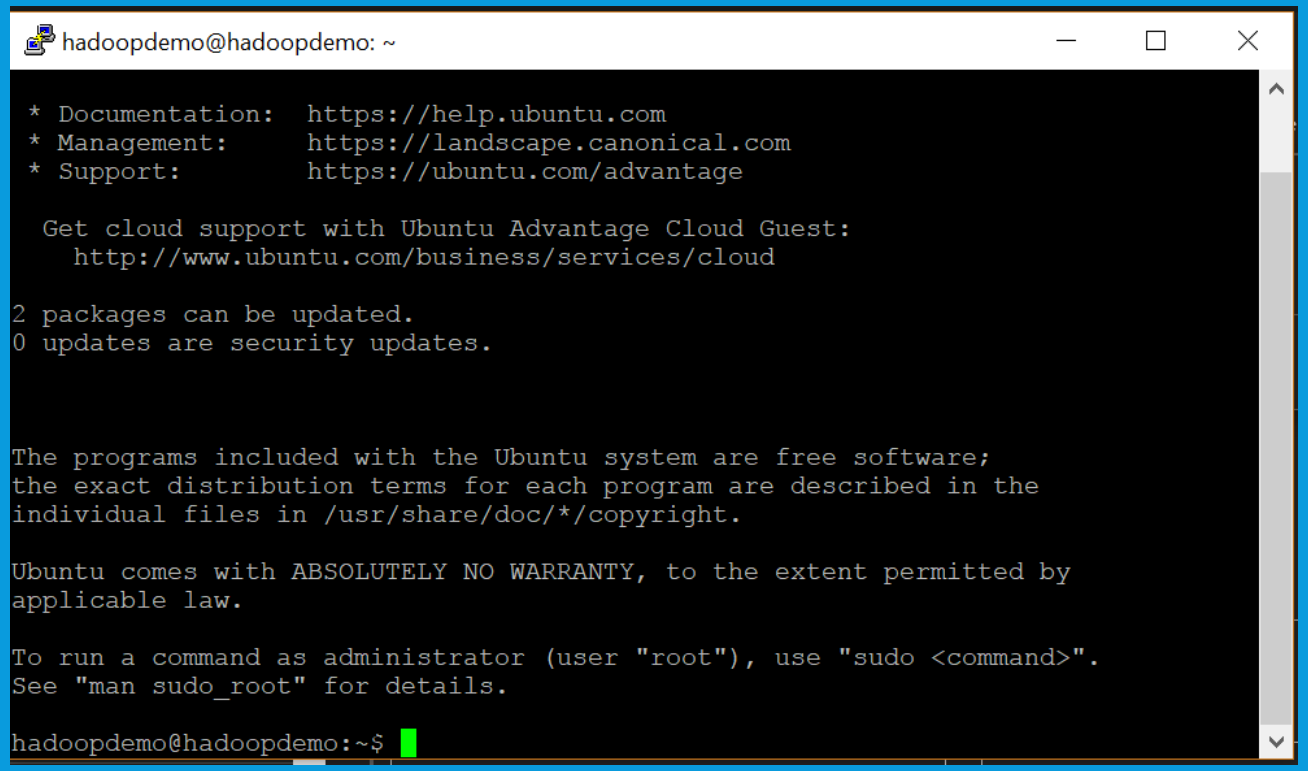
HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Connect to an instance from Windows using Putty**



HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- **Login**



HANDS-ON INSTALLING HADOOP

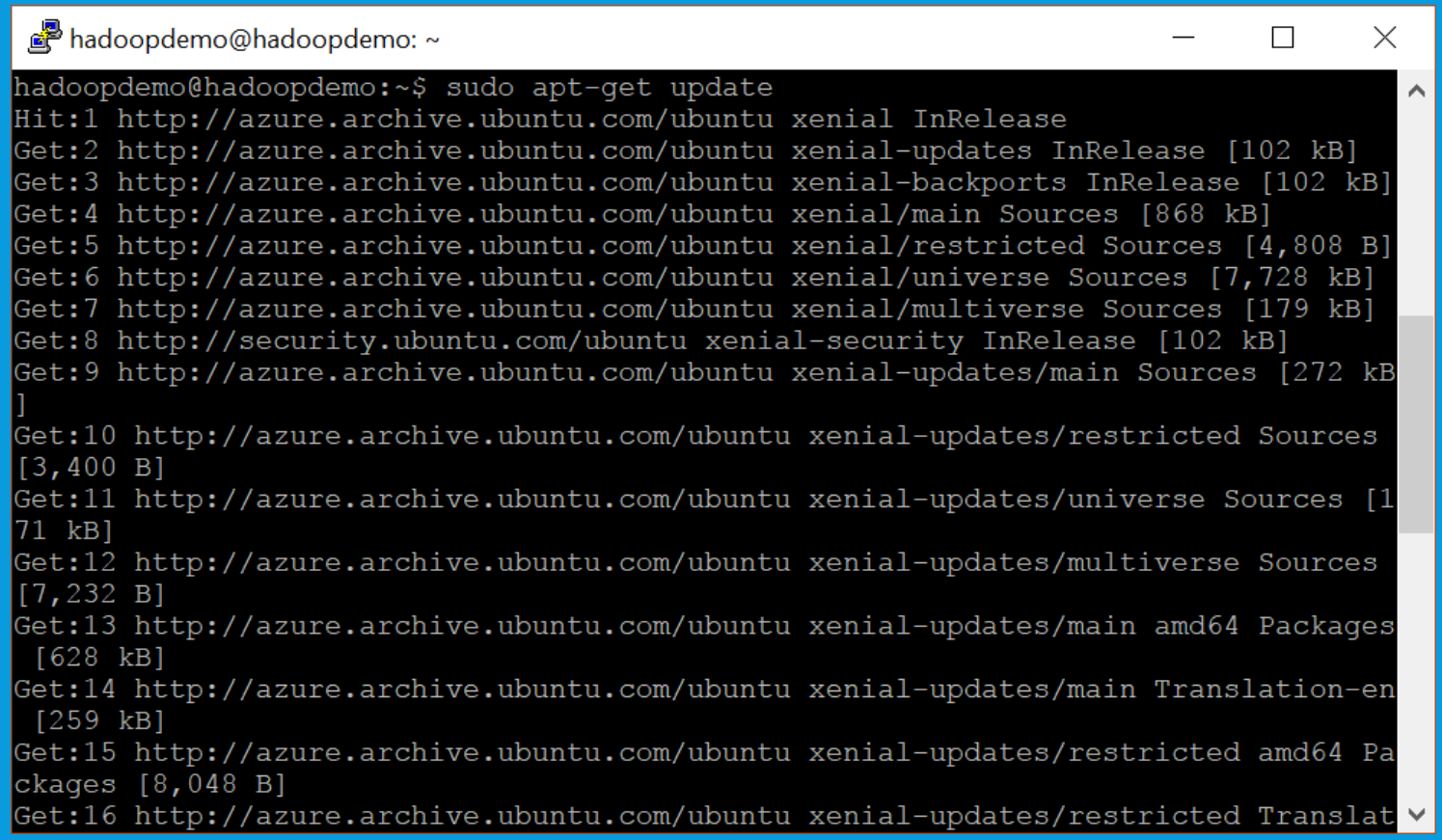
Installing Hadoop

1. Update System Software Repository
2. Configuring SSH
3. Installing Java
4. Download/Extract Hadoop
5. Installing Hadoop
6. Configure Hadoop
7. Formatting Namenode
8. Starting Hadoop
9. Accessing Hadoop Web Console
10. Stopping Hadoop

HANDS-ON INSTALLING HADOOP

1. Update System Software Repository

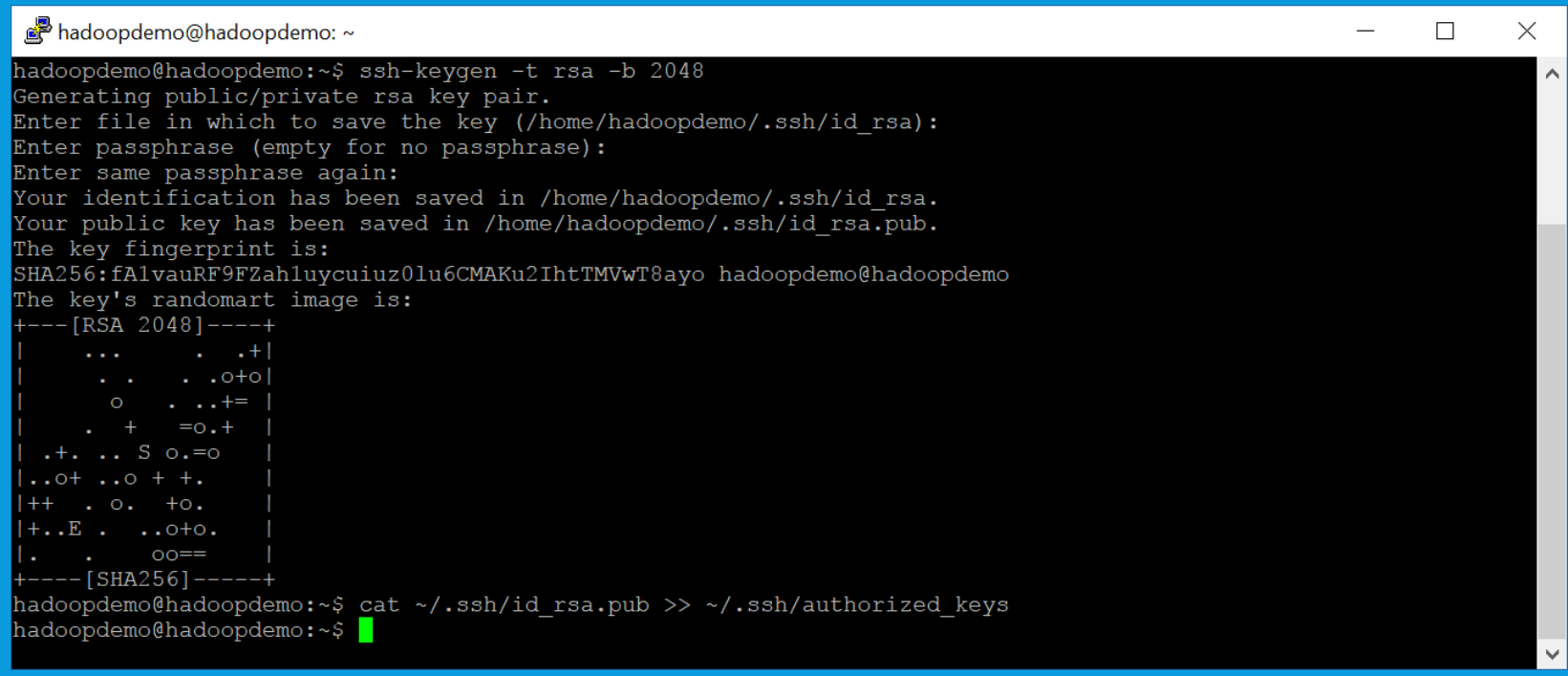
- *$ sudo apt-get update*



HANDS-ON INSTALLING HADOOP

2. Configuring SSH

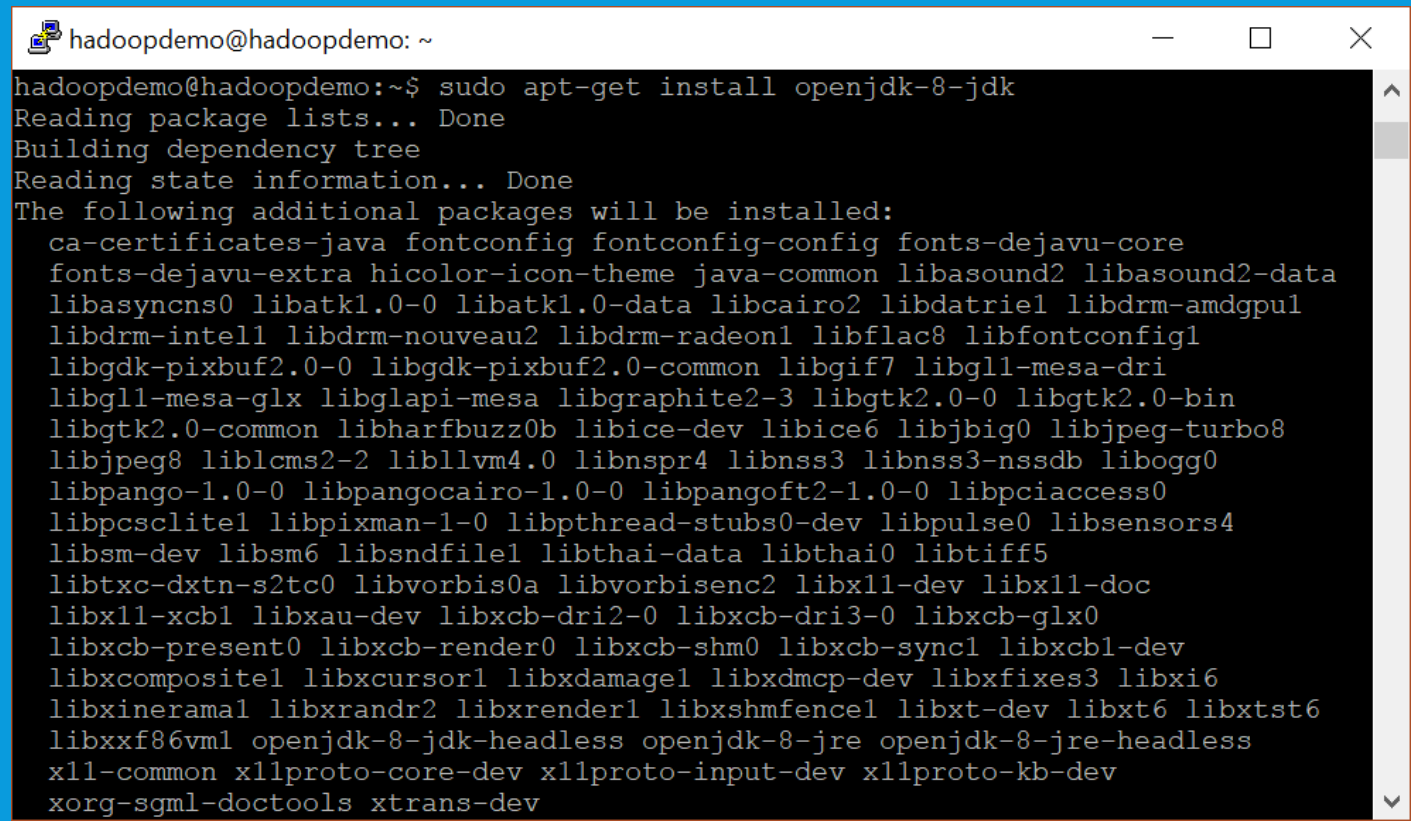
- *ssh-keygen -t rsa -b 2048*
- *$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys*



HANDS-ON INSTALLING HADOOP

3. Installing Java

$ sudo apt-get install openjdk-8-jdk

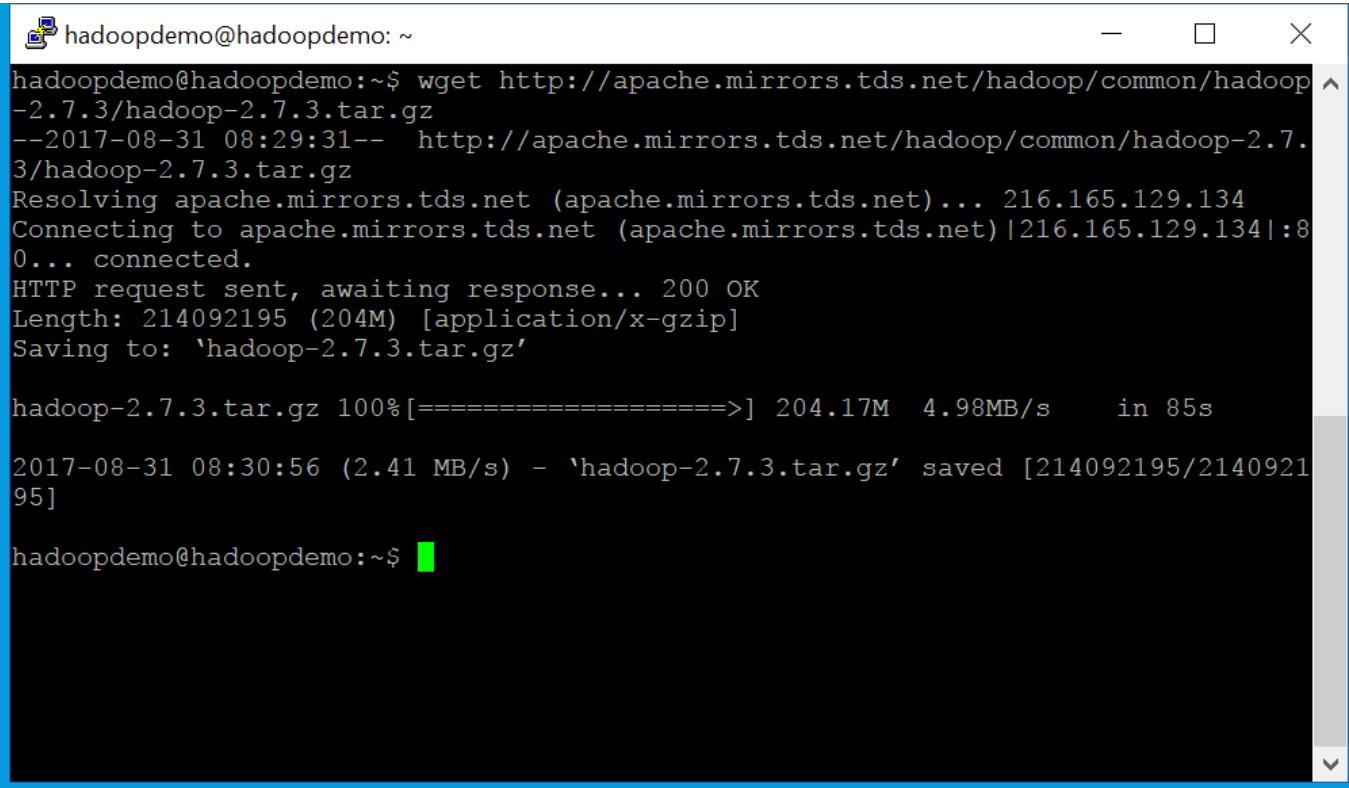


HANDS-ON INSTALLING HADOOP

4. Download/Extract Hadoop

- Download Apache Hadoop

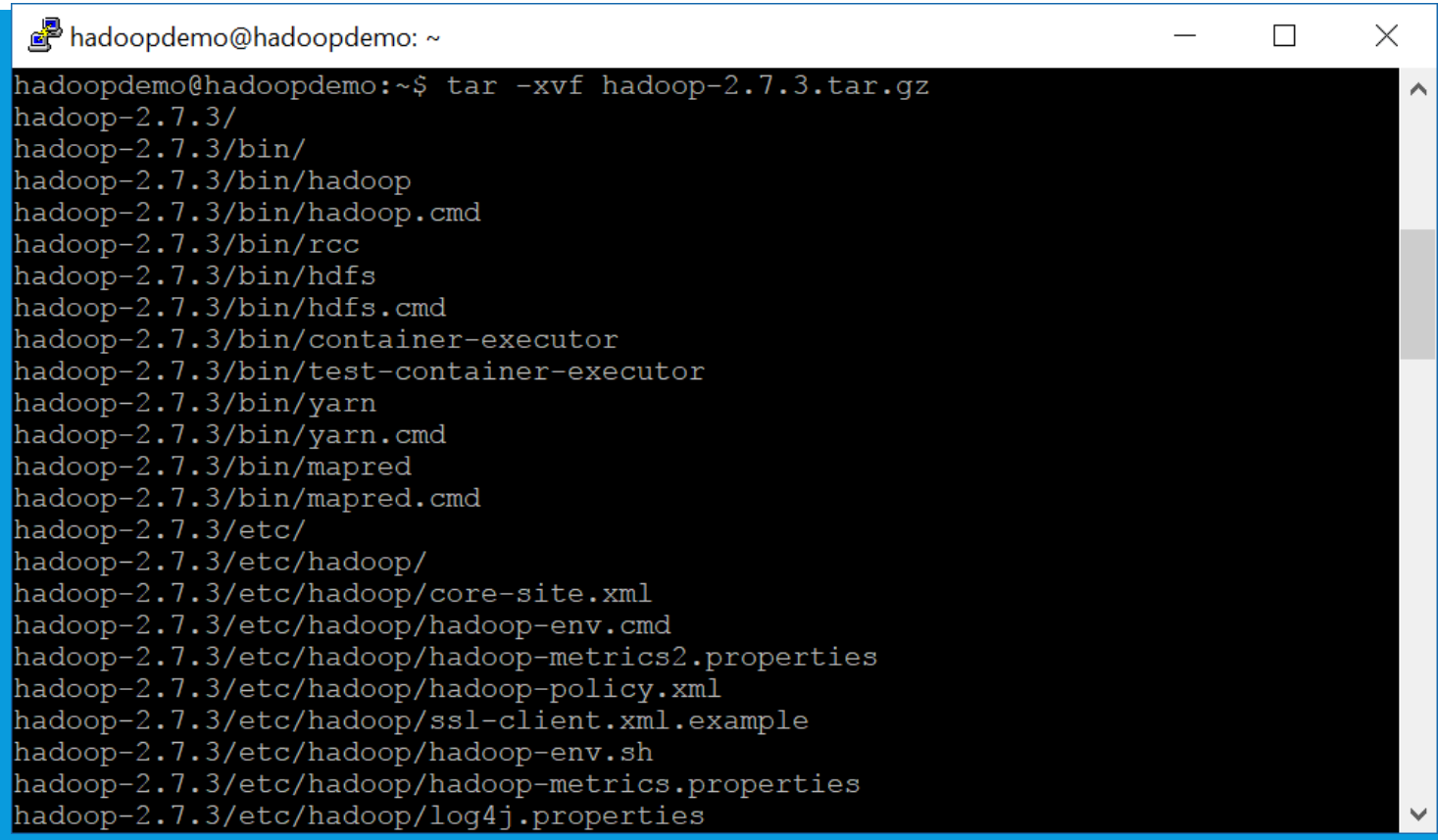
```
$ wget
http://apache.mirrors.tds.net/
hadoop/common/hadoop-
2.7.3/hadoop-2.7.3.tar.gz
```



HANDS-ON INSTALLING HADOOP

- **4. Download/Extract Hadoop**
- Extract Hadoop

$ tar -xvf hadoop-2.7.3.tar.gz

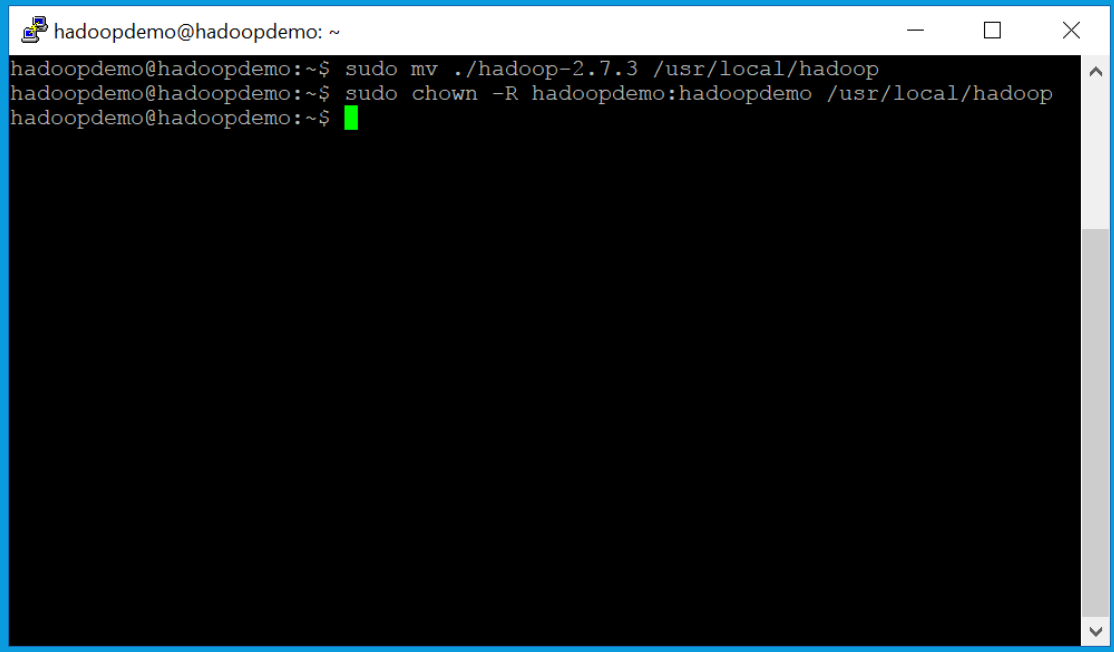


HANDS-ON INSTALLING HADOOP

5. Installing Hadoop

- *Move folder hadoop-2.7.3 to usr/local/hadoop*

```
$ sudo mv ./hadoop-2.7.3 /usr/local/hadoop
$ sudo chown -R user01:user01 /usr/local/hadoop
```



HANDS-ON INSTALLING HADOOP

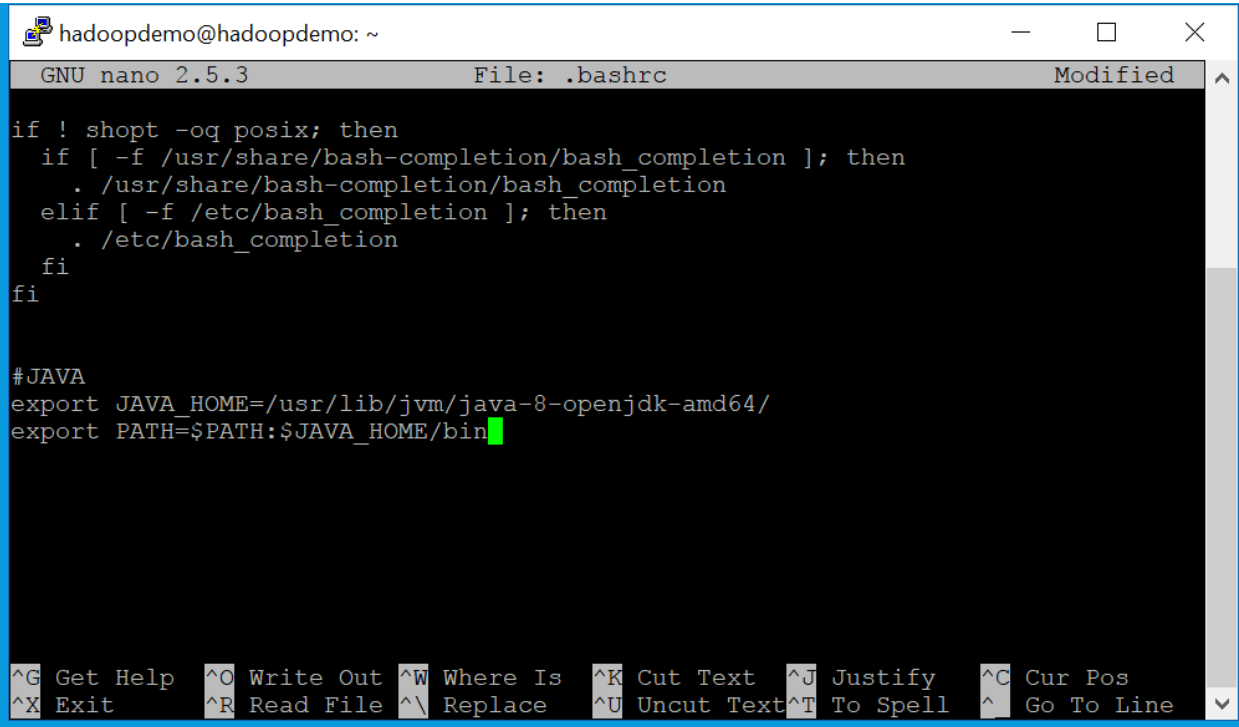
5.Installing Hadoop

- Set startup path and configuration bash shell

  $ sudo nano .bashrc

- Configuration java path

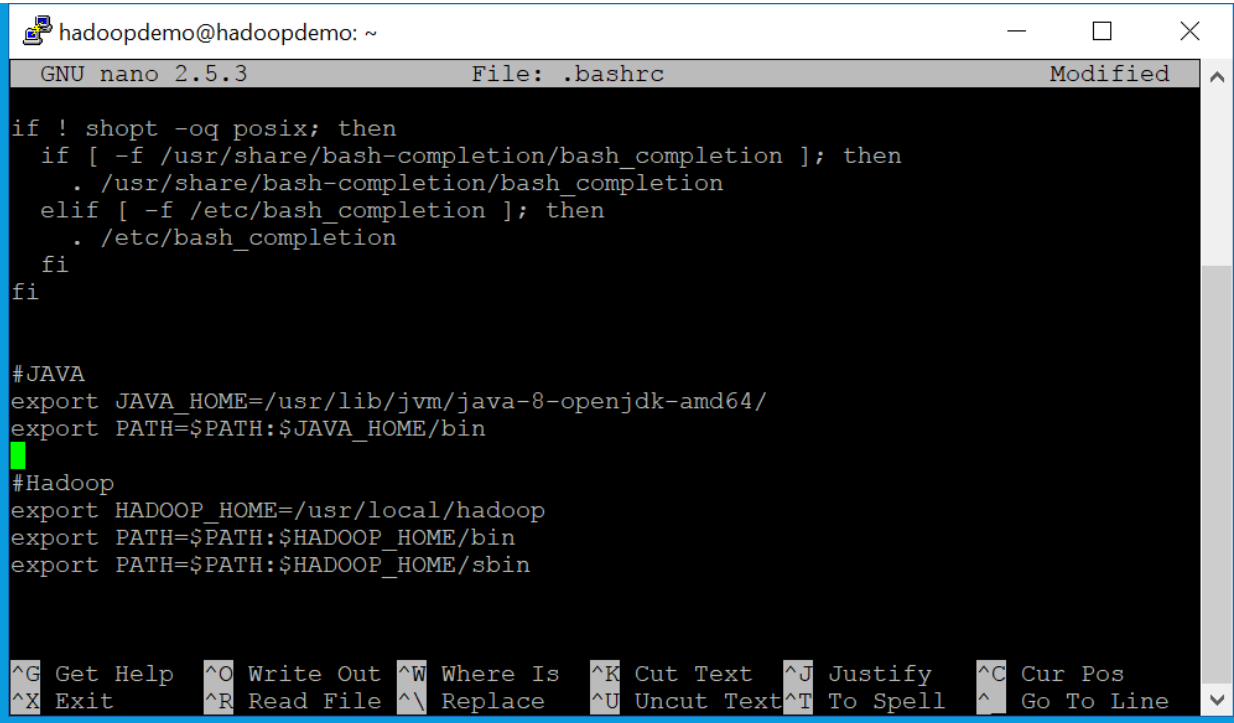
```
#JAVA

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/

export PATH=$PATH:$JAVA_HOME/bin
```



HANDS-ON INSTALLING HADOOP

- **6. Configure Hadoop**

```
#Hadoop
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
```

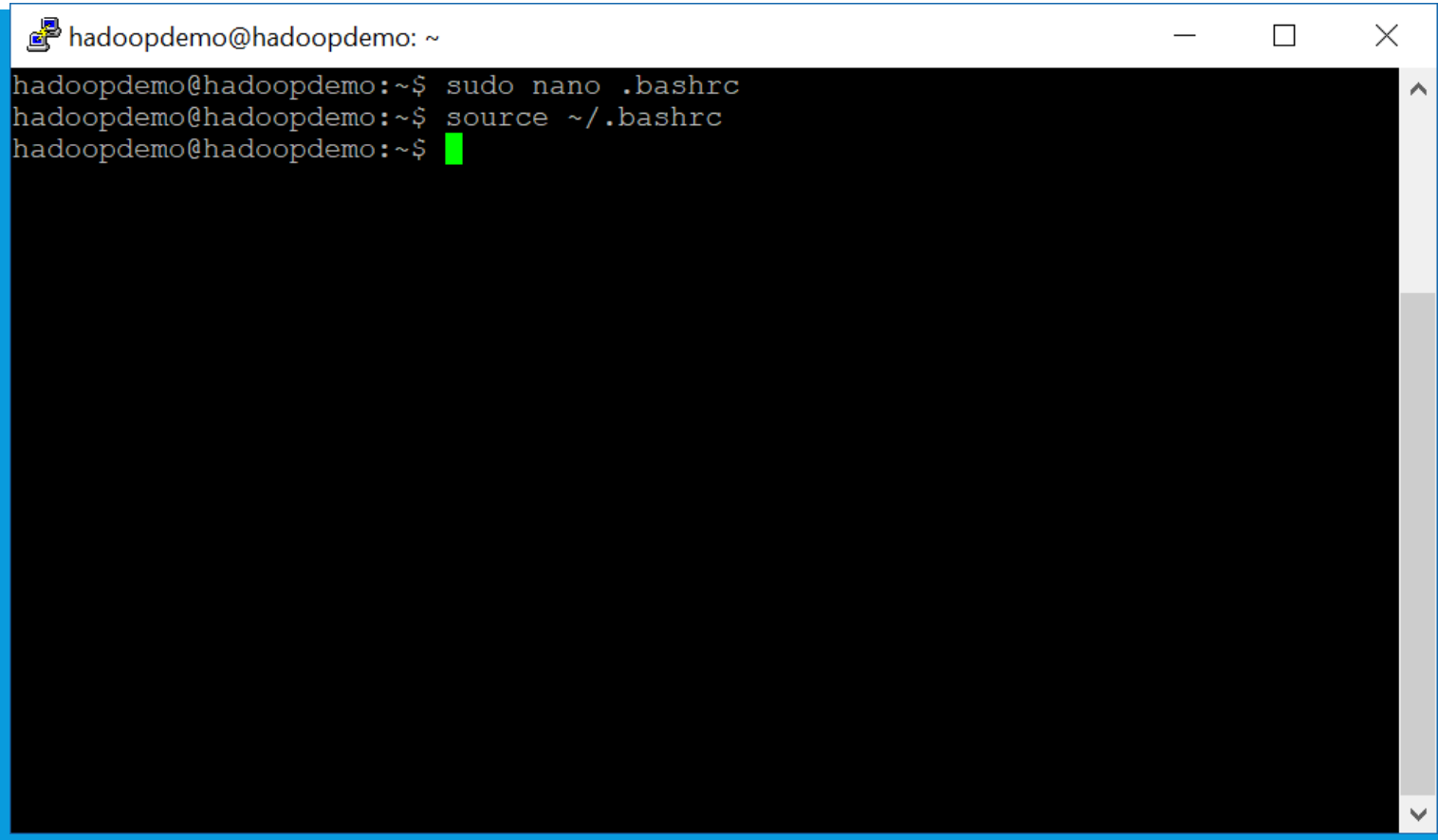


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Execute environment variables:

$ source ~/.bashrc

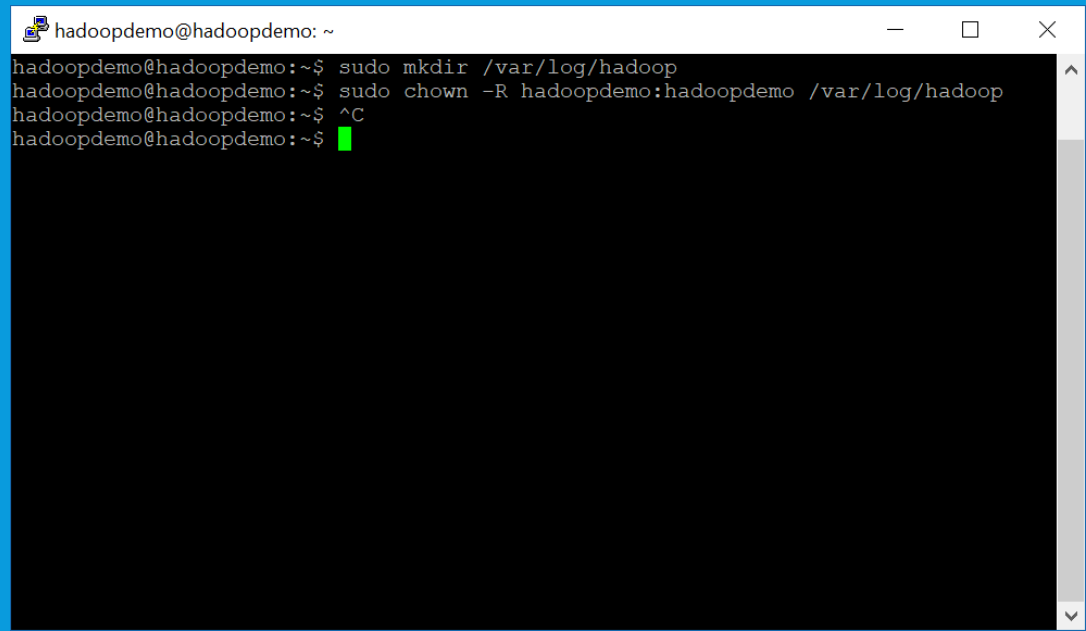


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Create hadoop log path

```
$ sudo mkdir /var/log/hadoop
$ sudo chown -R hadoopdemo:hadoopdemo /var/log/hadoop
```

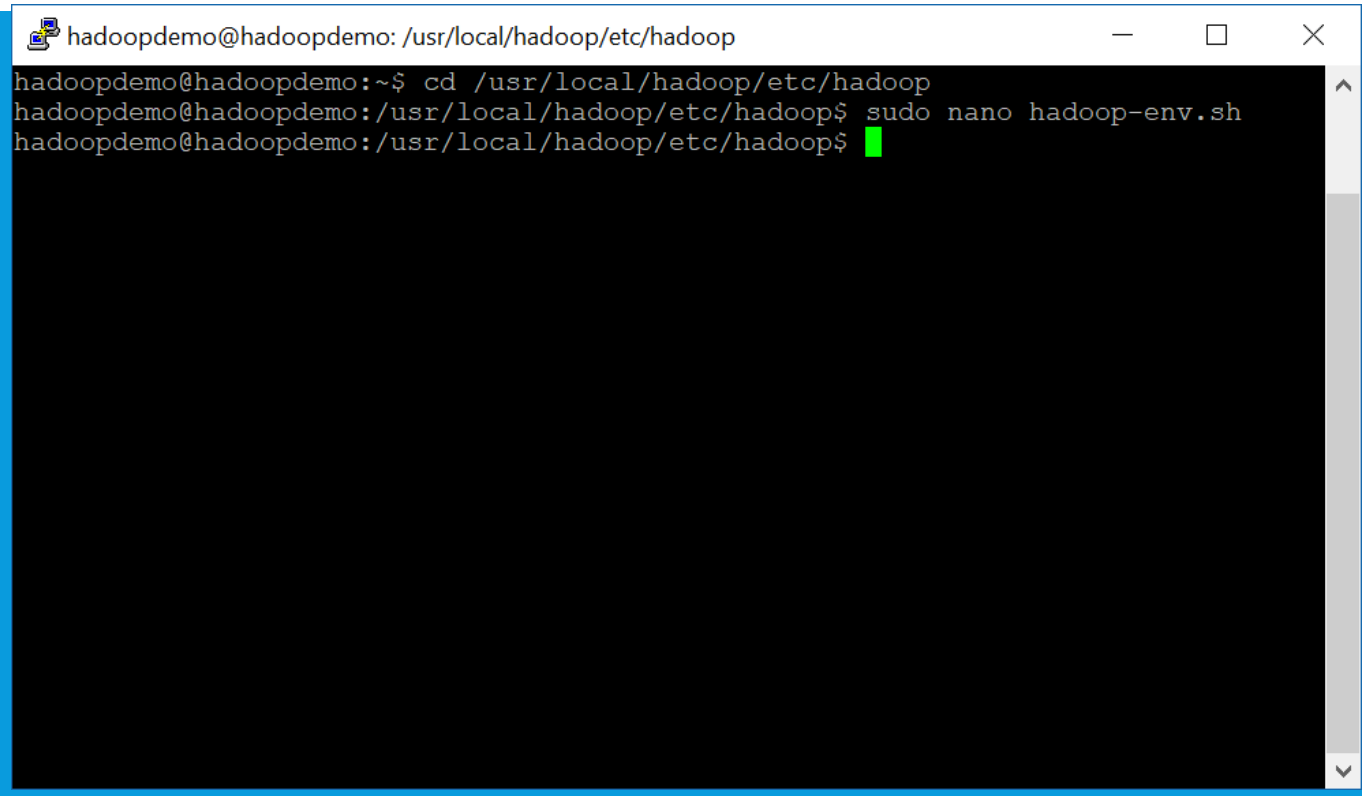


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Edit hadoop shell script

```
$ cd /usr/local/hadoop/etc/hadoop
$ sudo nano hadoop-env.sh
```

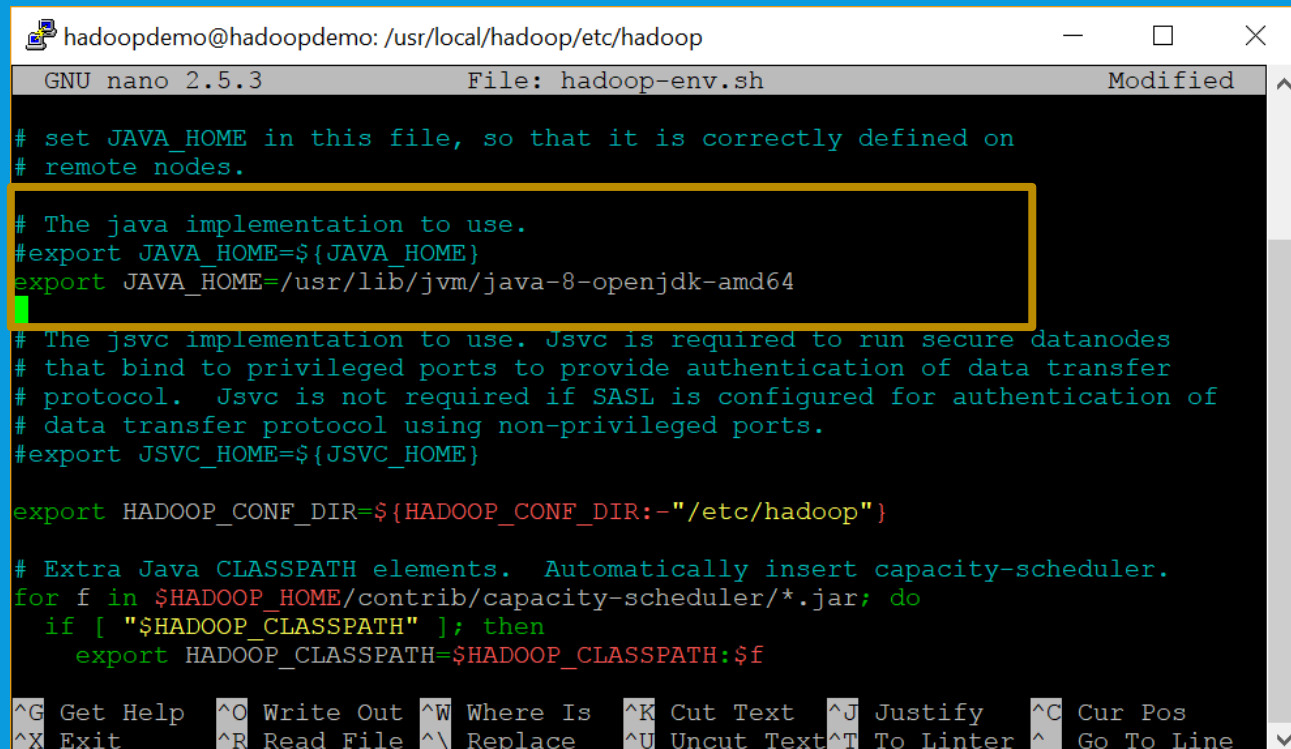


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Configuration JAVA_HOME

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

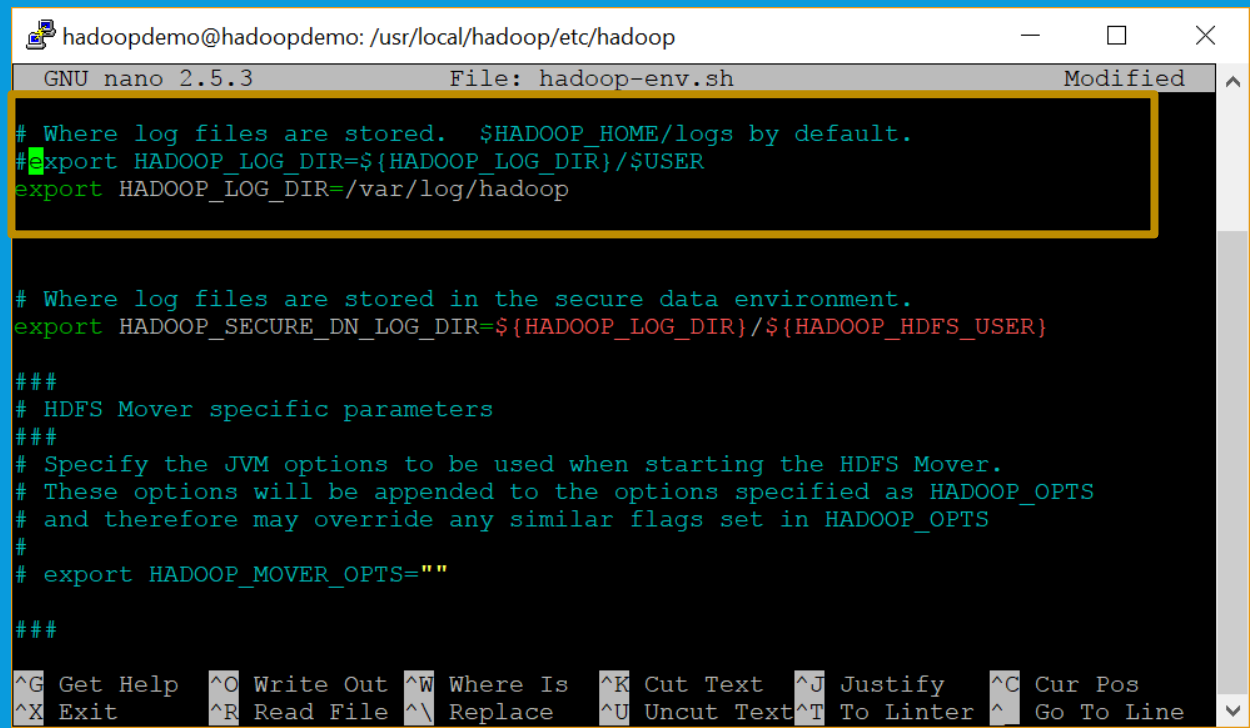


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Configuration hadoop log

```
export HADOOP_LOG_DIR=/var/log/hadoop
```



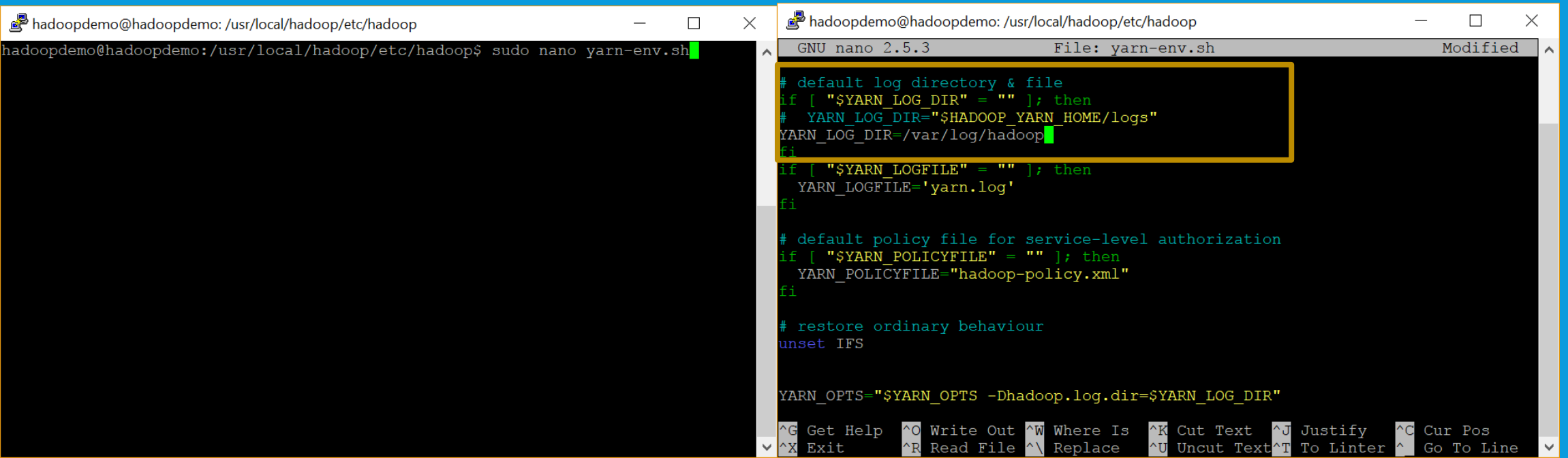
HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Edit Yarn shell script – log location to another directory

$ sudo nano yarn-env.sh

YARN_LOG_DIR=/var/log/hadoop

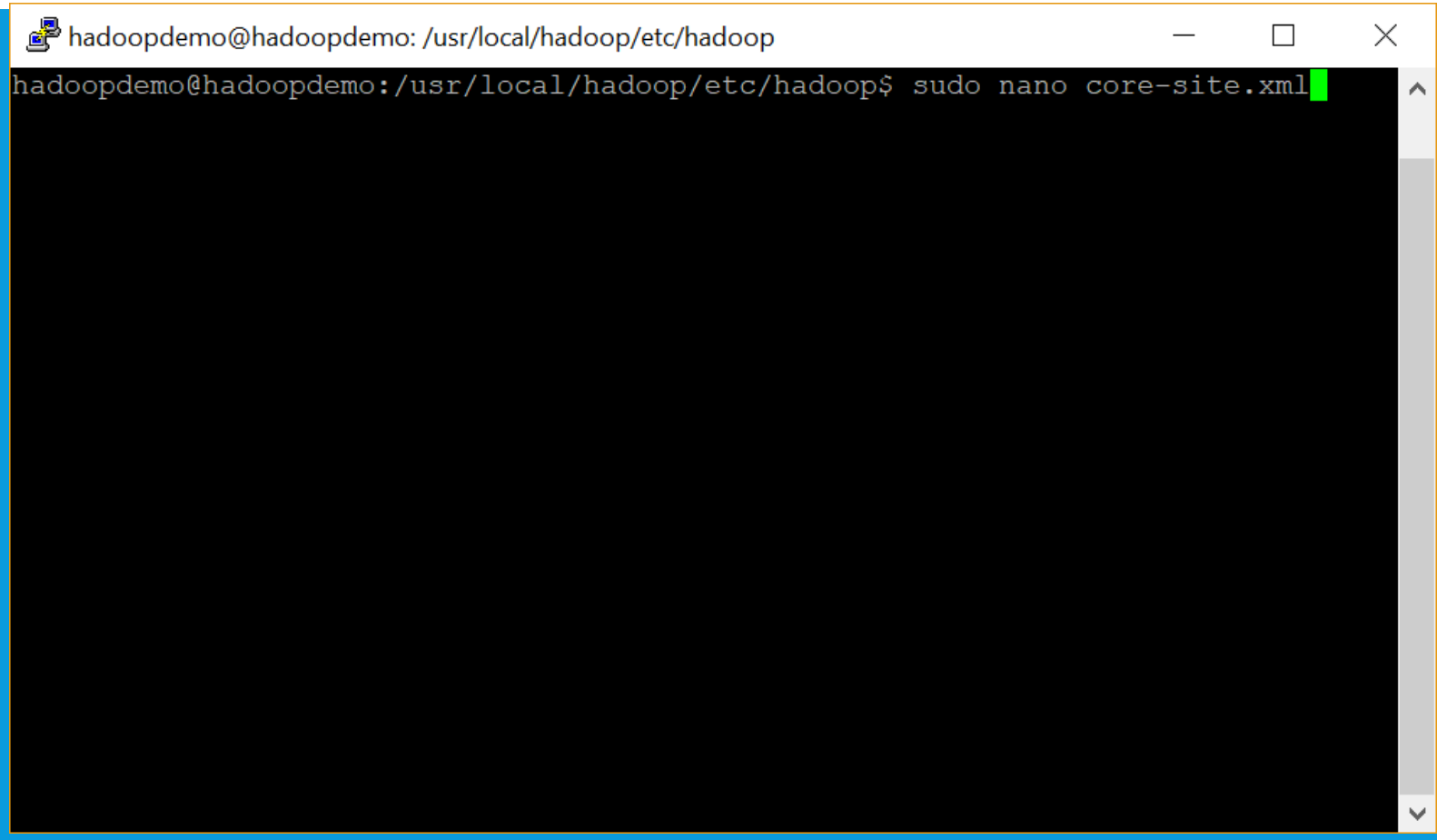


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Edit hadoop - core-site.xml

```
$ sudo nano core-site.xml
```

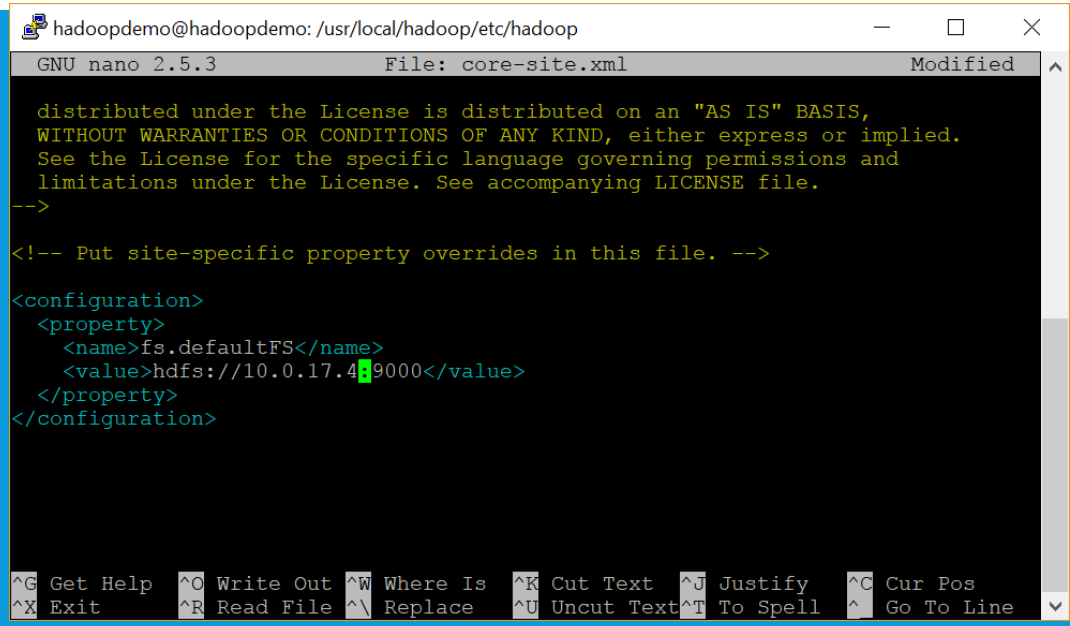


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- My Ip is 10.0.17.4

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://10.0.17.4:9000</value>
  </property>
</configuration>
```

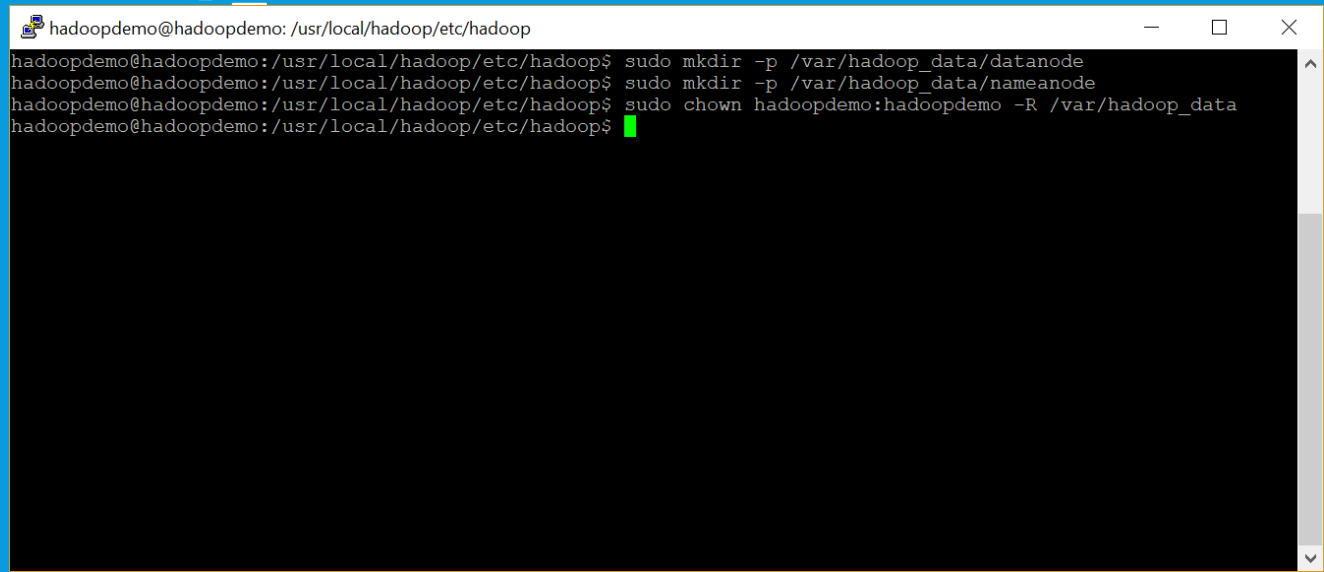


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Create namenode and datanode path

```
$ sudo mkdir -p /var/hadoop_data/namenode

$ sudo mkdir -p /var/hadoop_data/datanode

$ sudo chown user01:user01 -R /var/hadoop_data
```

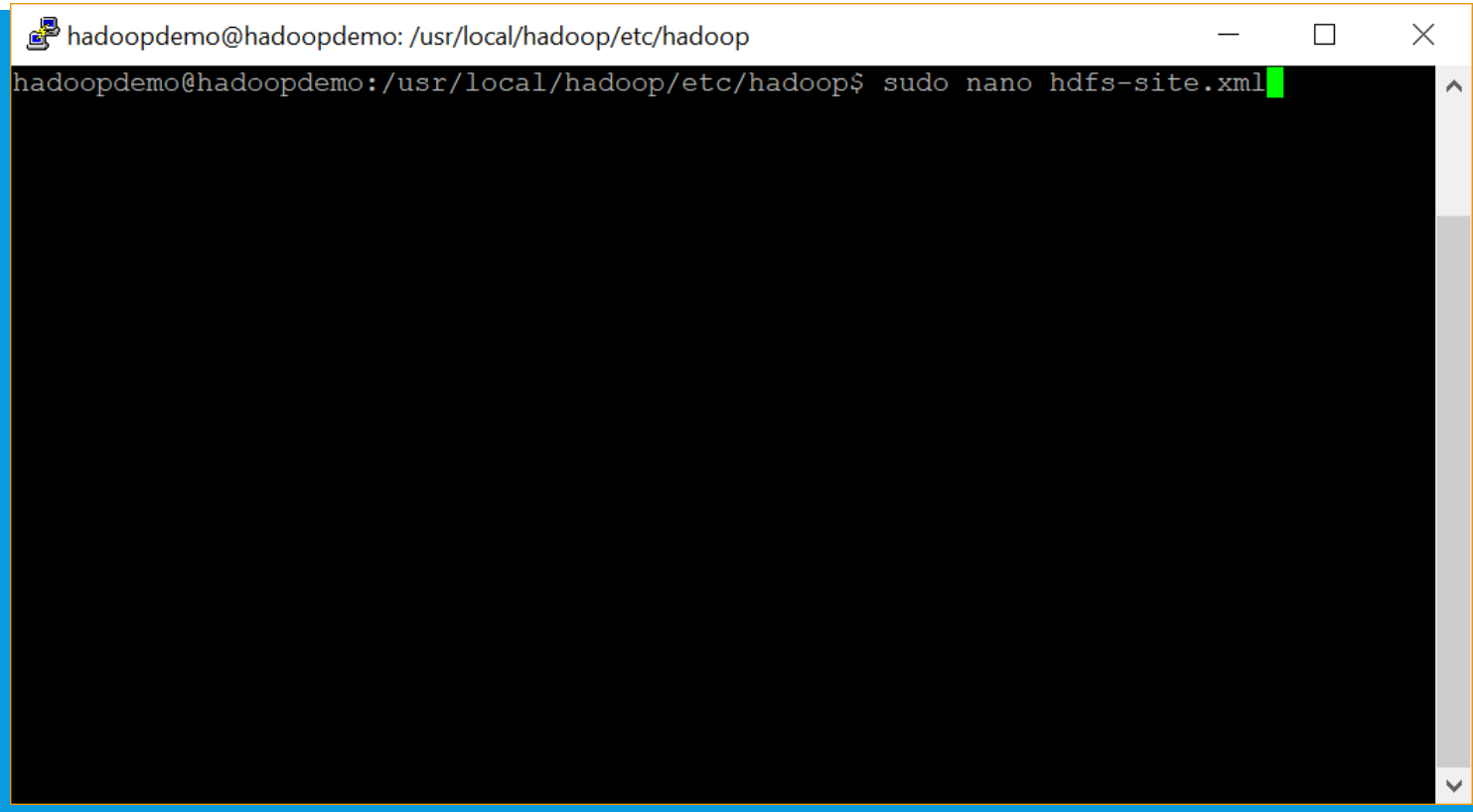


HANDS-ON INSTALLING HADOOP

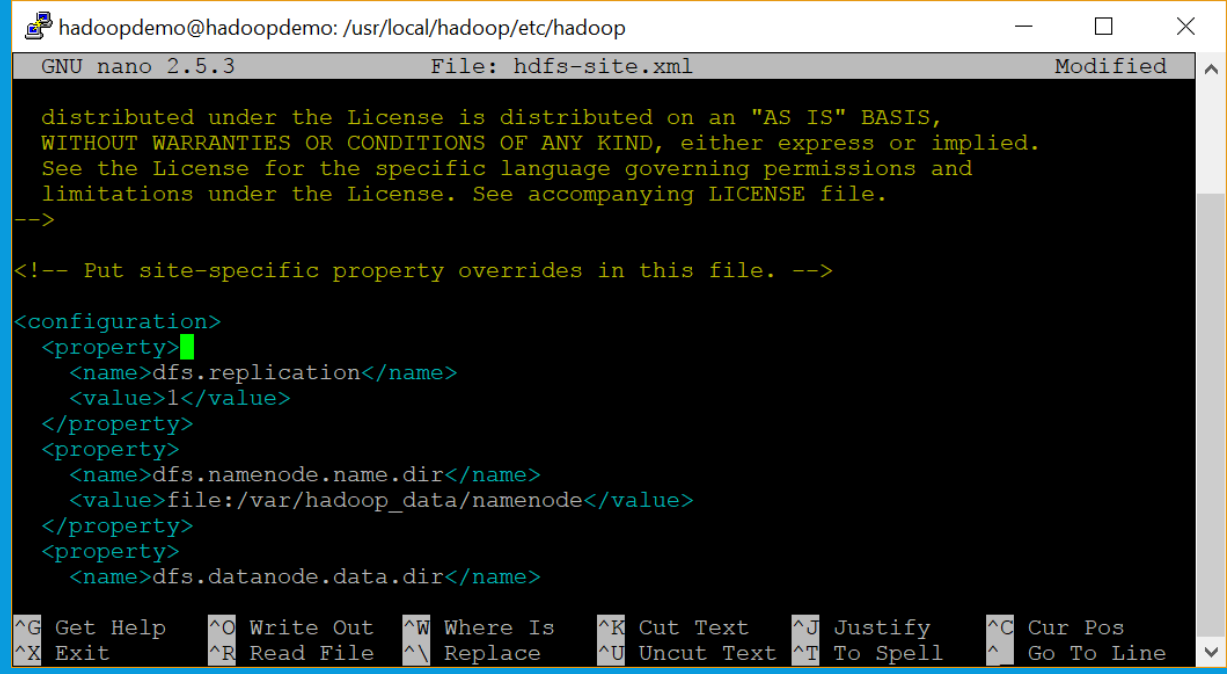
6. Configure Hadoop

- Edit hdfs-site.xml

$ sudo nano hdfs-site.xml



```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/var/hadoop_data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/var/hadoop_data/datanode</value>
  </property>
</configuration>
```

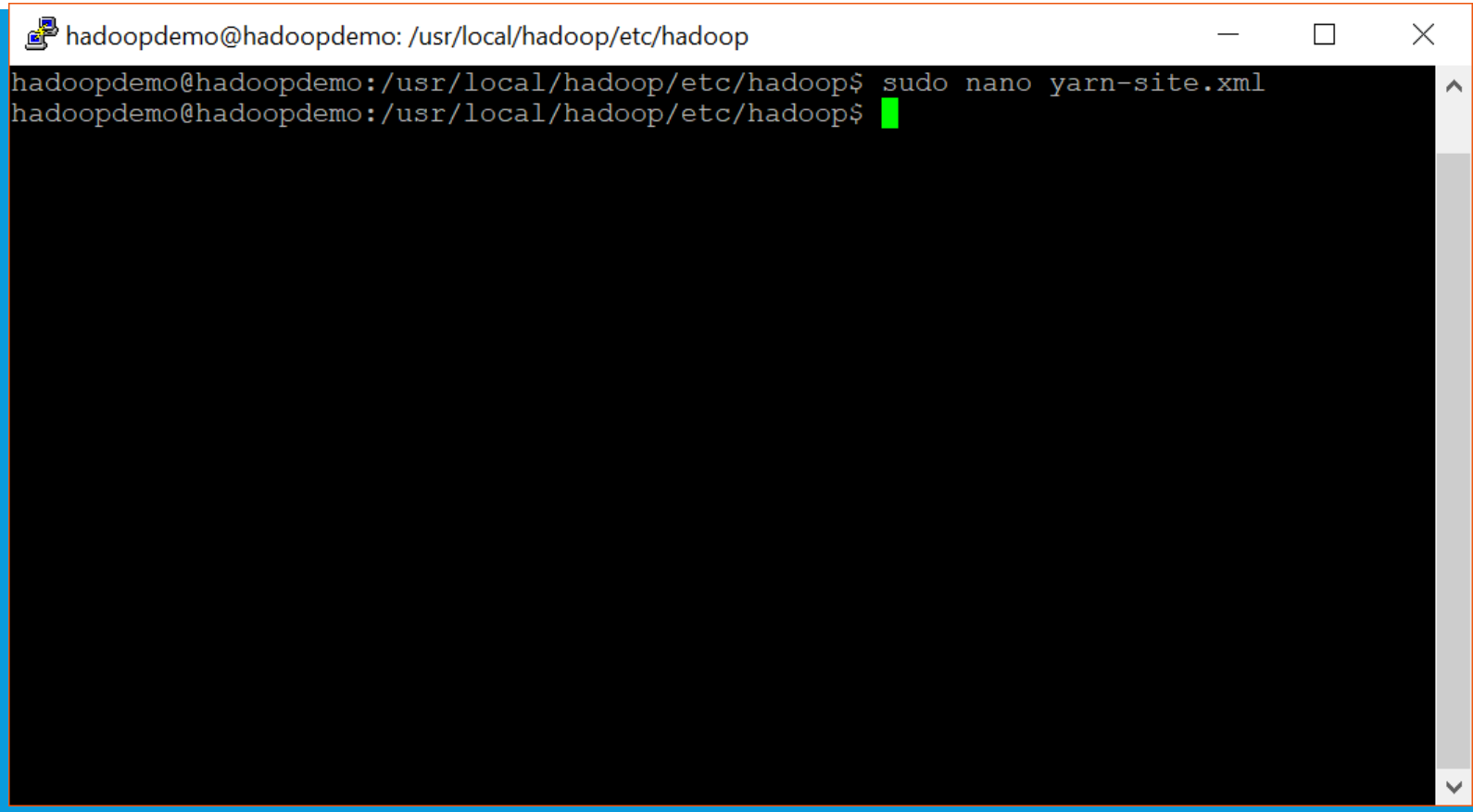


HANDS-ON INSTALLING HADOOP

6. Configure Hadoop

- Edit yarn-site.xml

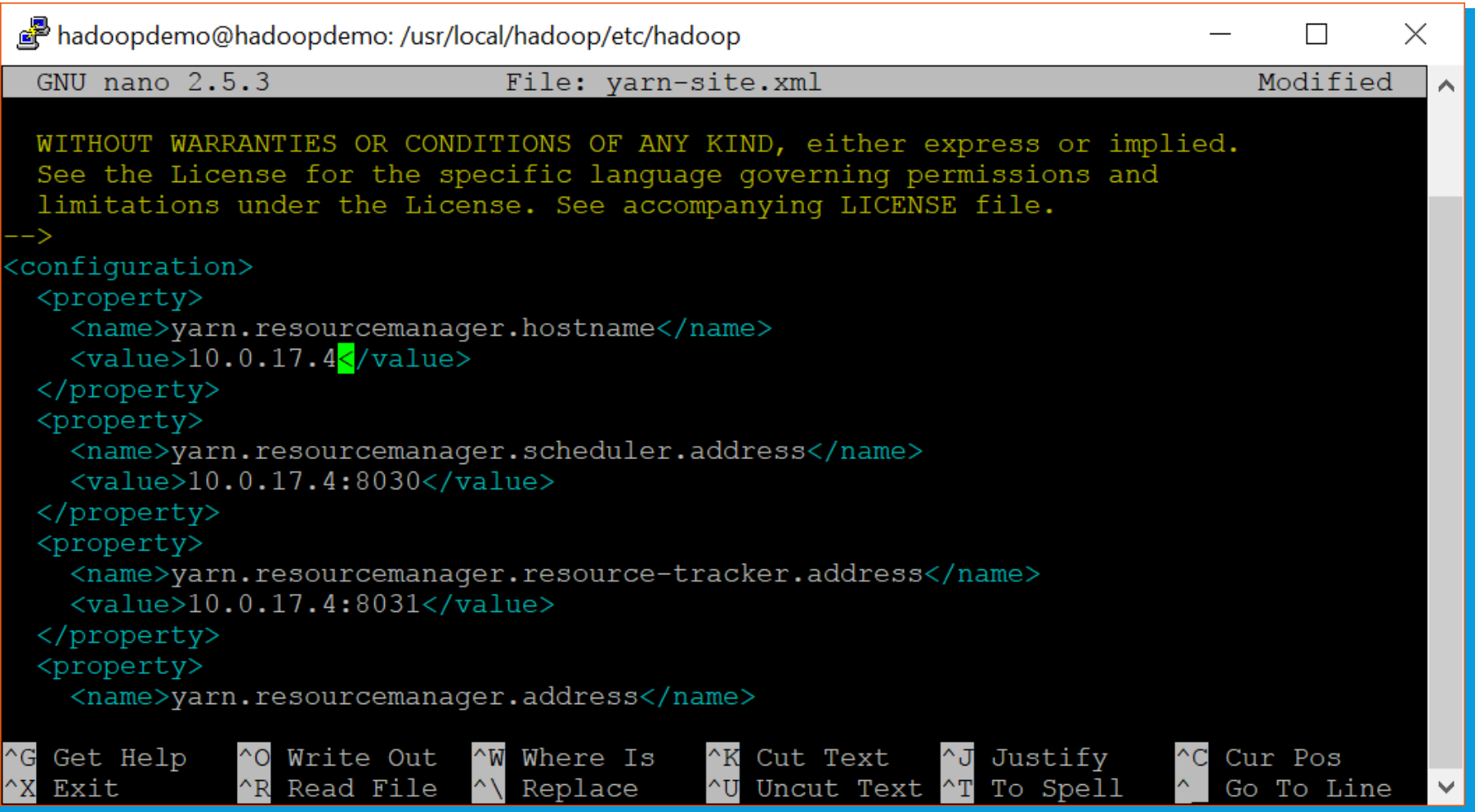
```
$ sudo nano yarn-site.xml
```



6. CONFIGURING HADOOP

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>10.0.17.4</value>
  </property>
  <property>
    <name>yarn.resourcemanager.scheduler.address</name>
    <value>10.0.17.4:8030</value>
  </property>
  <property>
    <name>yarn.resourcemanager.resource-tracker.address</name>
    <value>10.0.17.4:8031</value>
  </property>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>10.0.17.4:8032</value>
  </property>
  <property>
    <name>yarn.resourcemanager.admin.address</name>
    <value>10.0.17.4:8033</value>
  </property>
  <property>
    <name>yarn.resourcemanager.webapp.address</name>
    <value>10.0.17.4:8088</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-
services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

HANDS-ON INSTALLING HADOOP

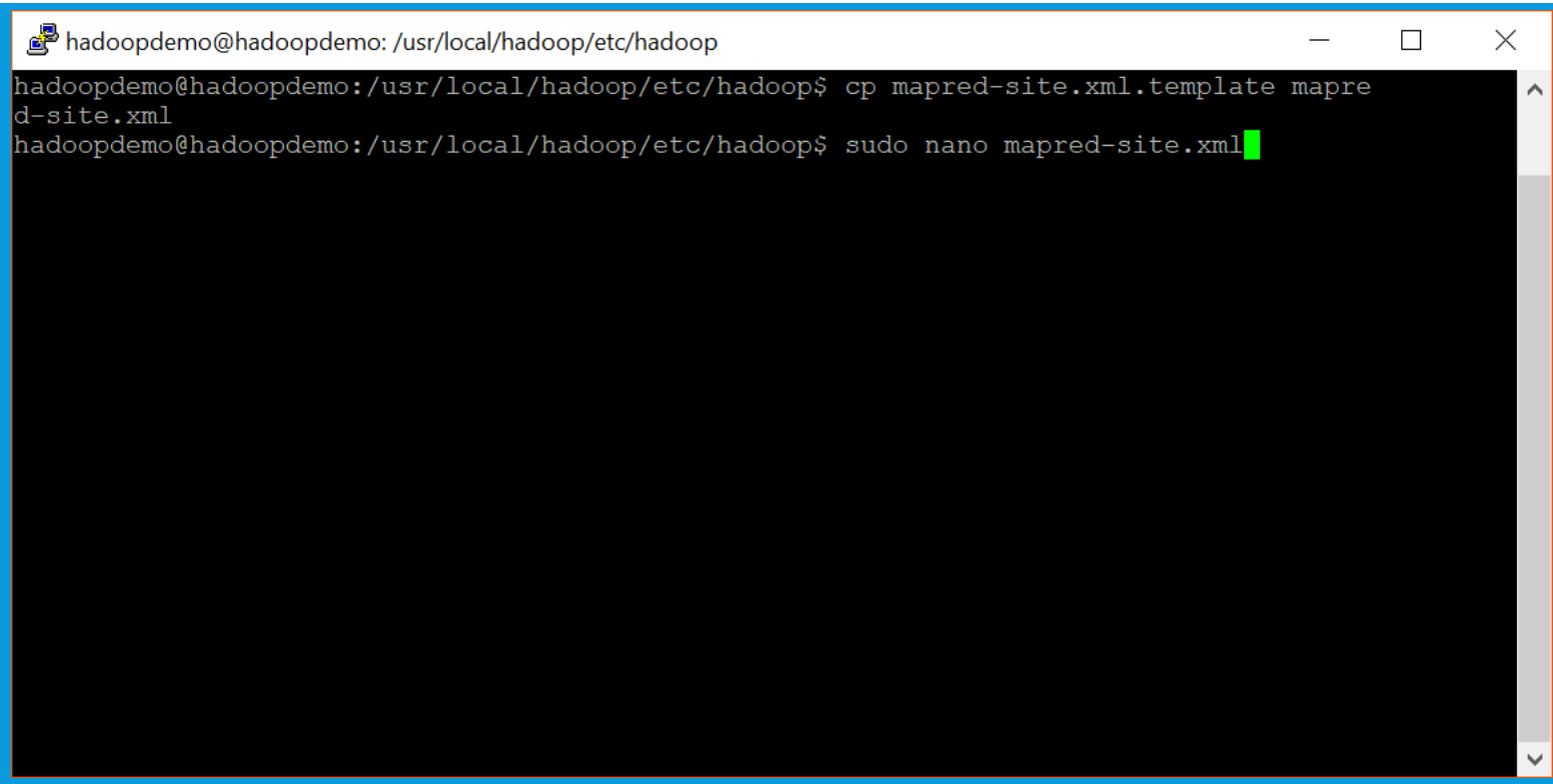


HANDS-ON INSTALLING HADOOP

6.Configure Hadoop

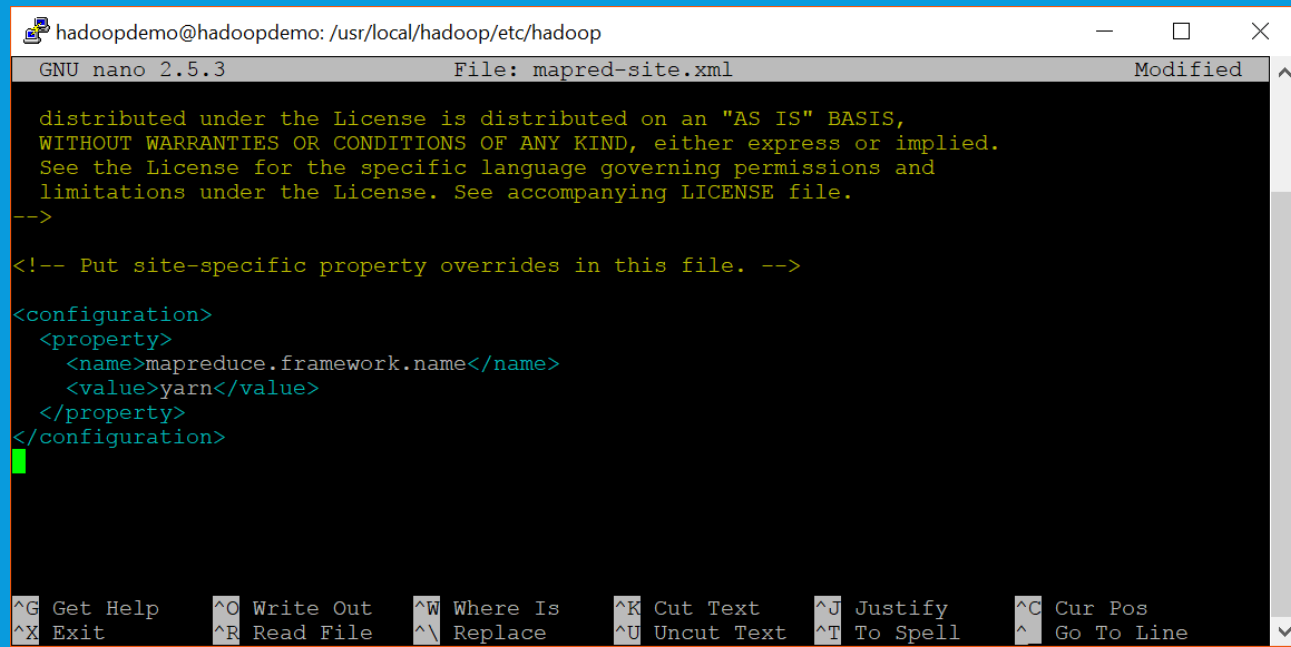
- Edit mapred-site.xml

$ sudo nano mapred-site.xml



HANDS-ON INSTALLING HADOOP

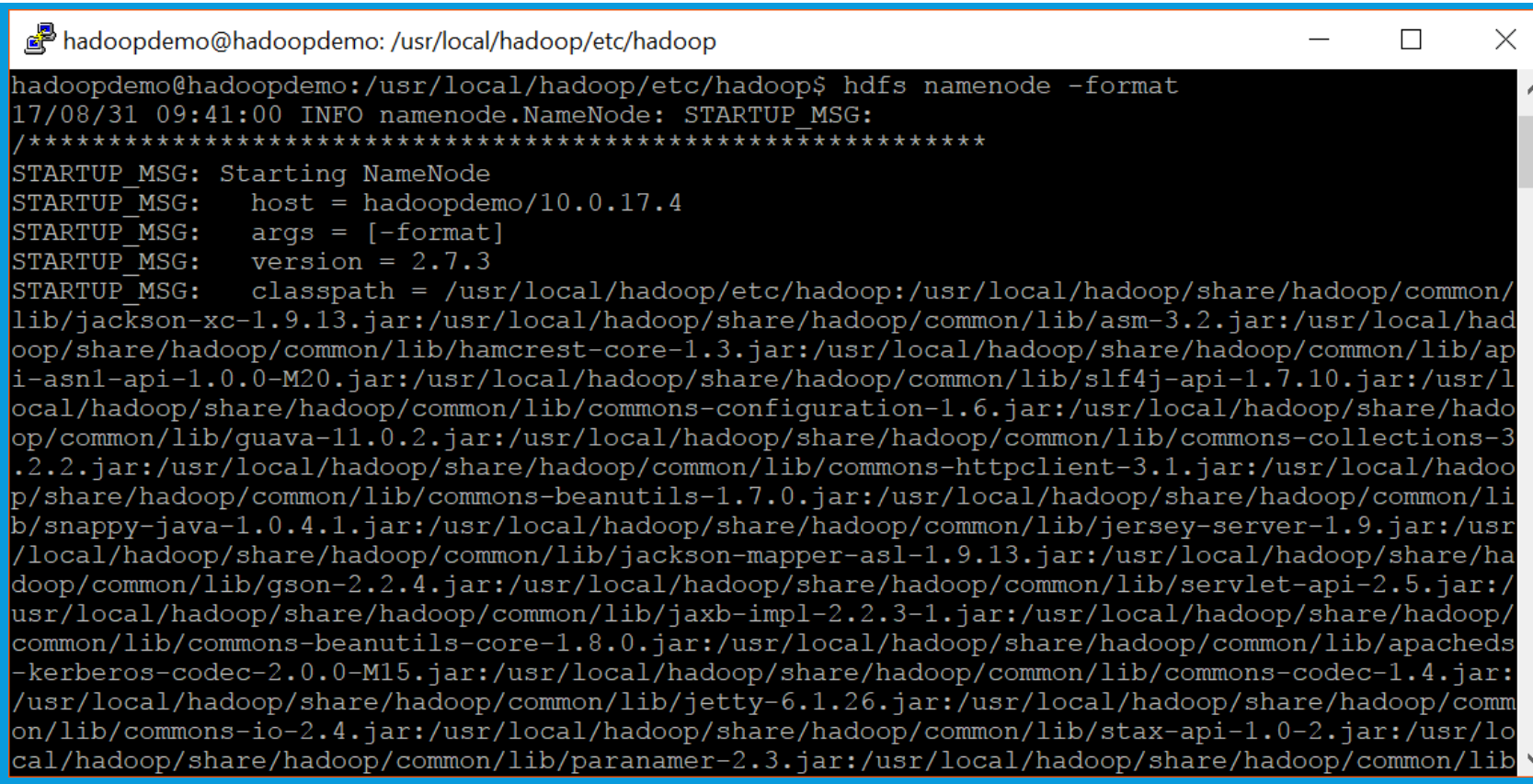
```
<configuration>
  <property>

<name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```



HANDS-ON INSTALLING HADOOP

7.Formatting Namenode

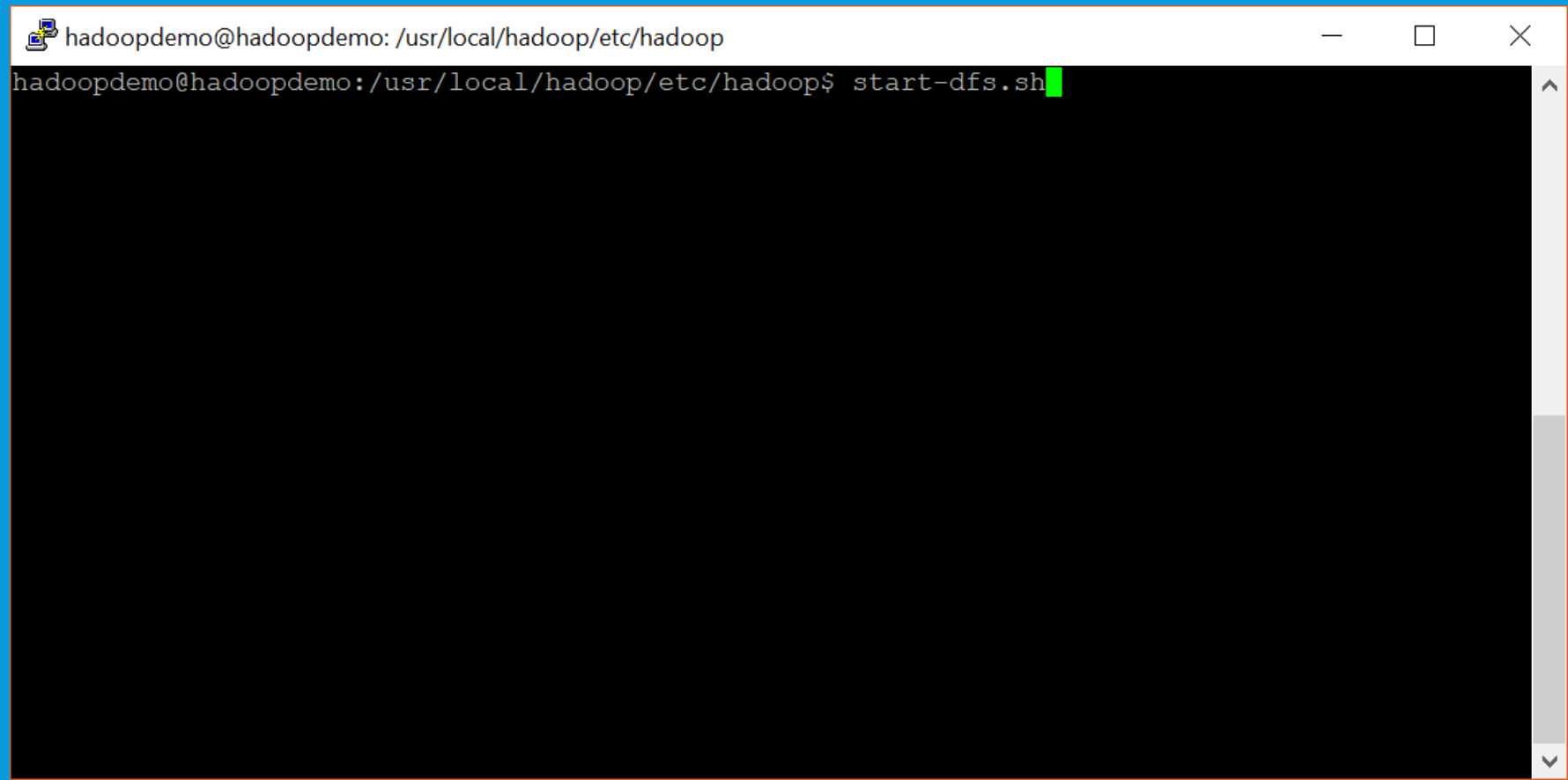
$ hdfs namenode -format



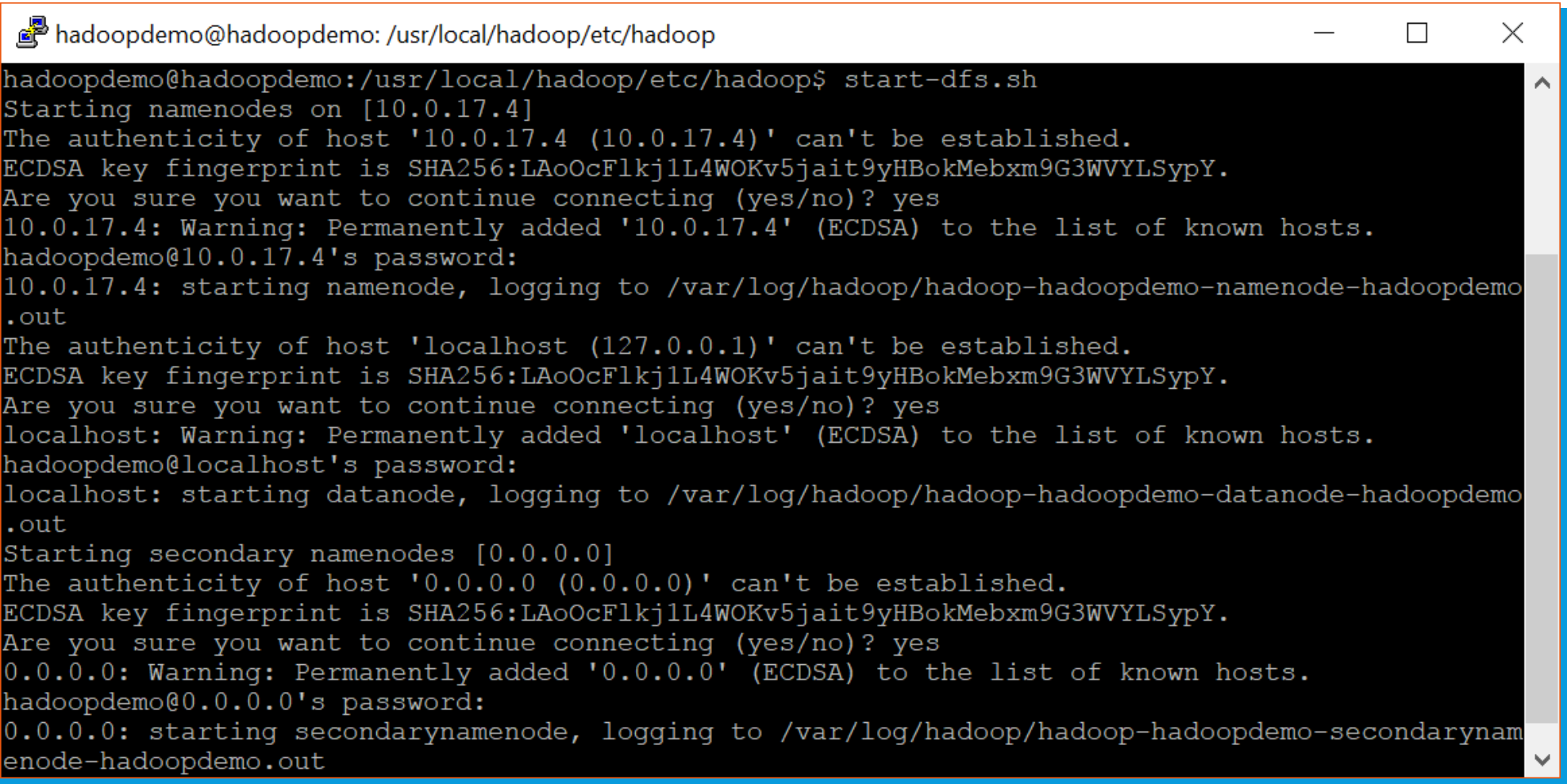
HANDS-ON INSTALLING HADOOP

- **8.Starting Namenode and Datanode**

```
$ start-dfs.sh
```



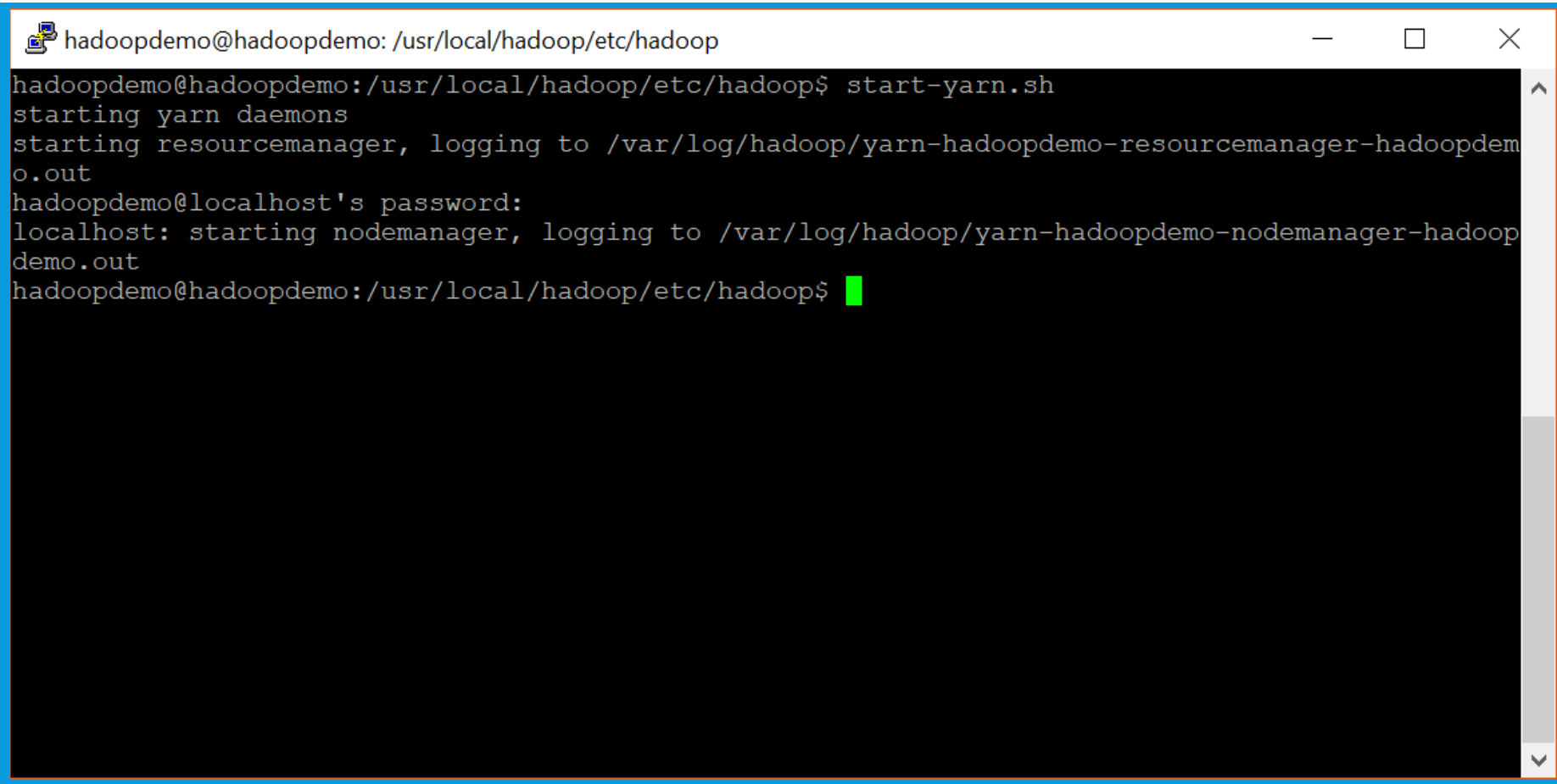
HANDS-ON INSTALLING HADOOP



HANDS-ON INSTALLING HADOOP

8.Starting Yarn

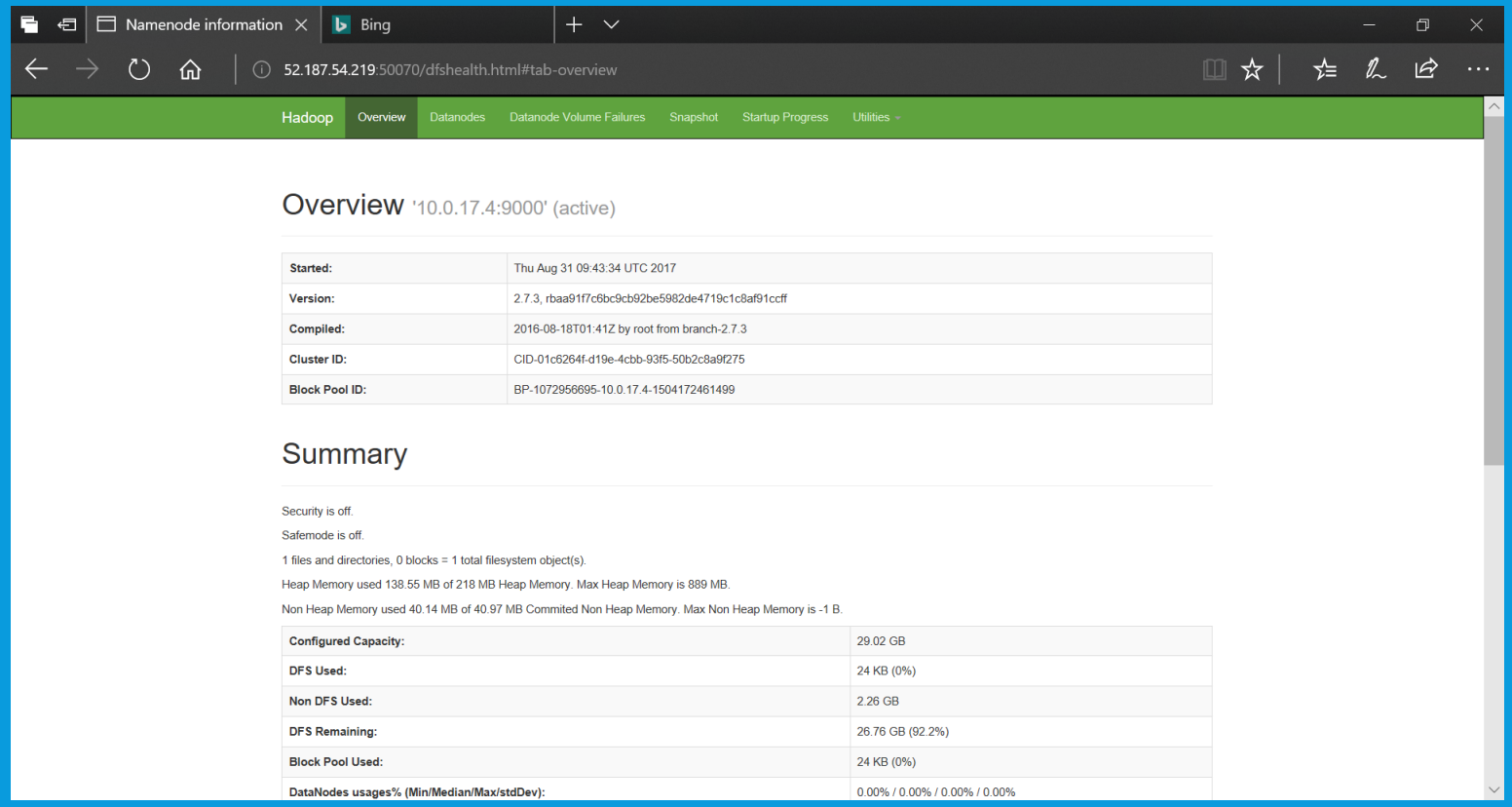
$ start-yarn.sh



HANDS-ON INSTALLING HADOOP

9. Accessing Hadoop Web Console

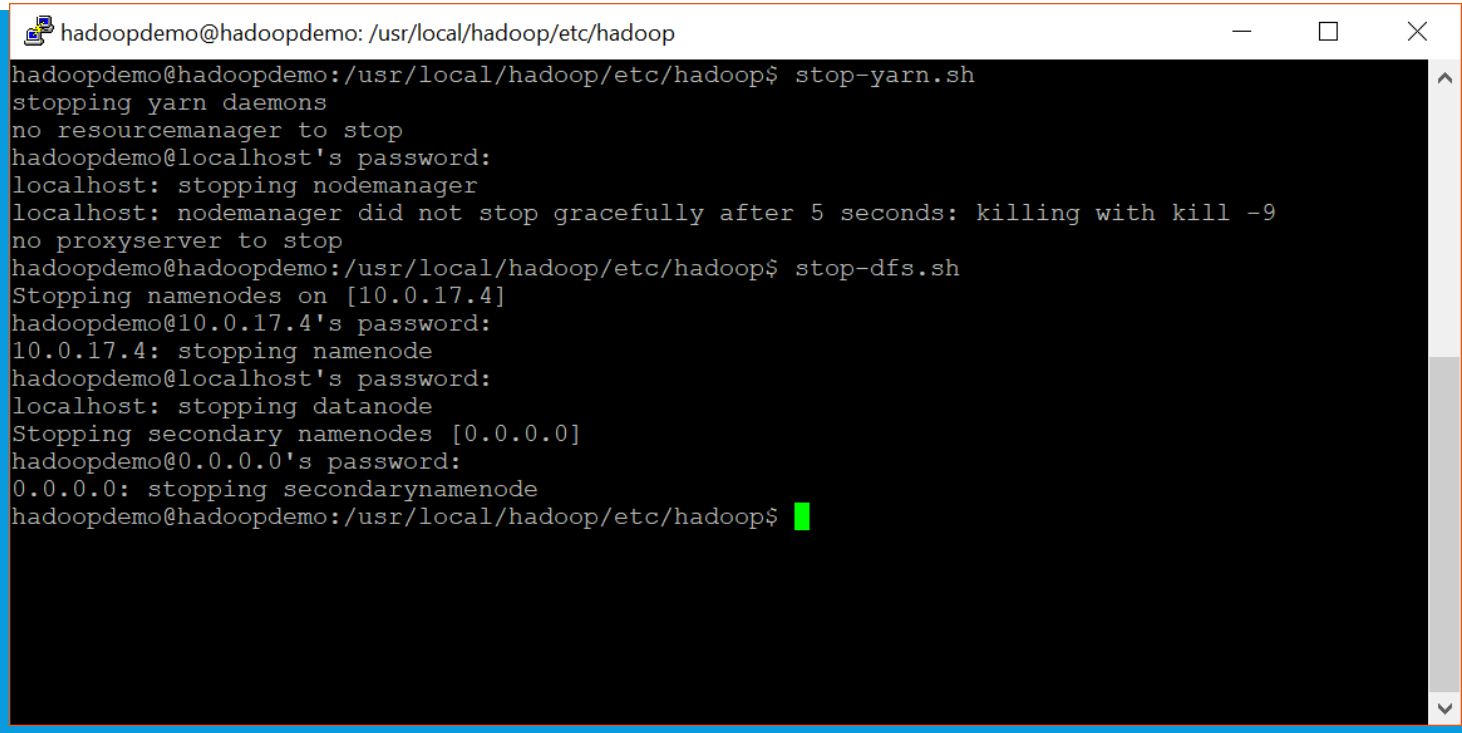
- My public ip is 52.187.54.219



HANDS-ON INSTALLING HADOOP

10.Stop Hadoop Yarn and HDFS

```
$ stop-yarn.sh
$ stop-dfs.sh
```



HANDS-ON: TRAVERSING, RETRIEVING DATA FROM HDFS

- **Create input and ouput folder on dfs**

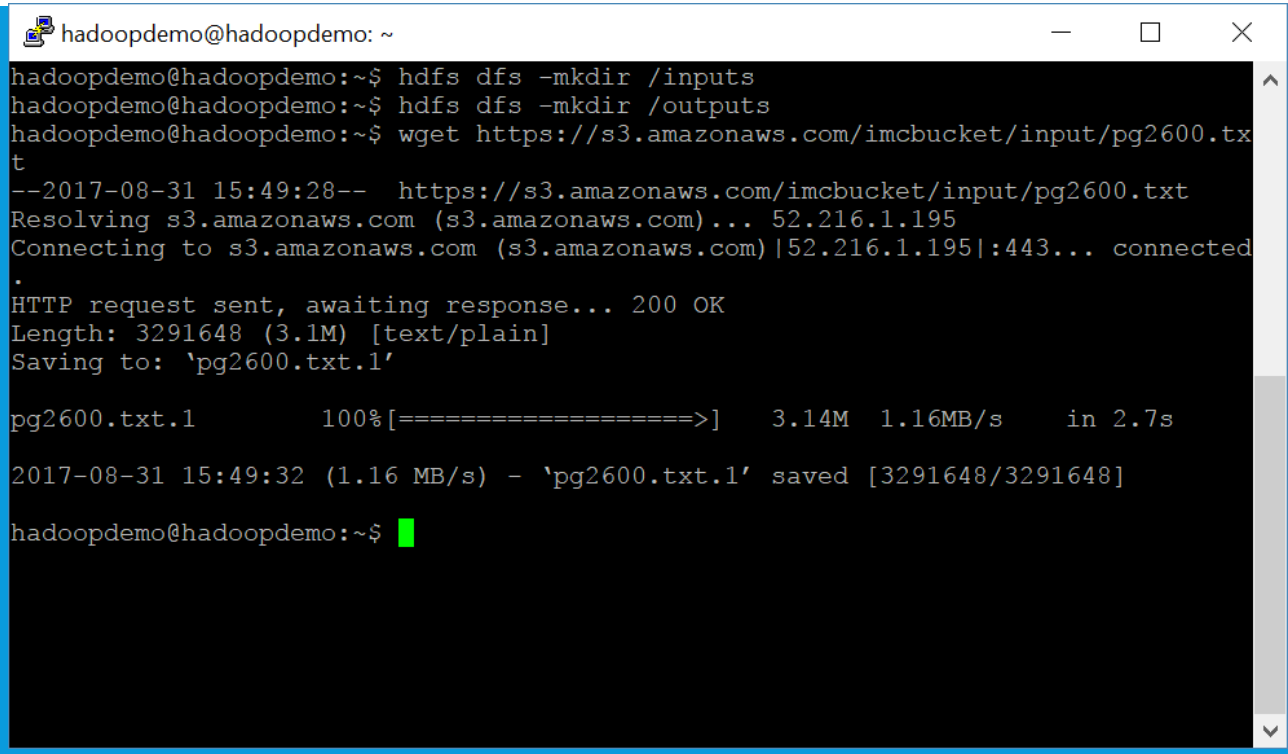
$ hdfs dfs -mkdir /inputs

$ hdfs dfs -mkdir /outputs

Get data file from internet

$ wget https://s3.amazonaws.com/imcbucket/input/pg2600.txt

$ hdfs dfs -copyFromLocal ~/pg2600.txt /inputs/



HANDS-ON: TRAVERSING, RETRIEVING DATA FROM HDFS

```
$ hdfs dfs -ls /inputs
$ hdfs dfs -cat /inputs/pg2600.txt
$ hdfs dfs -rm /inputs/pg2600.txt
```