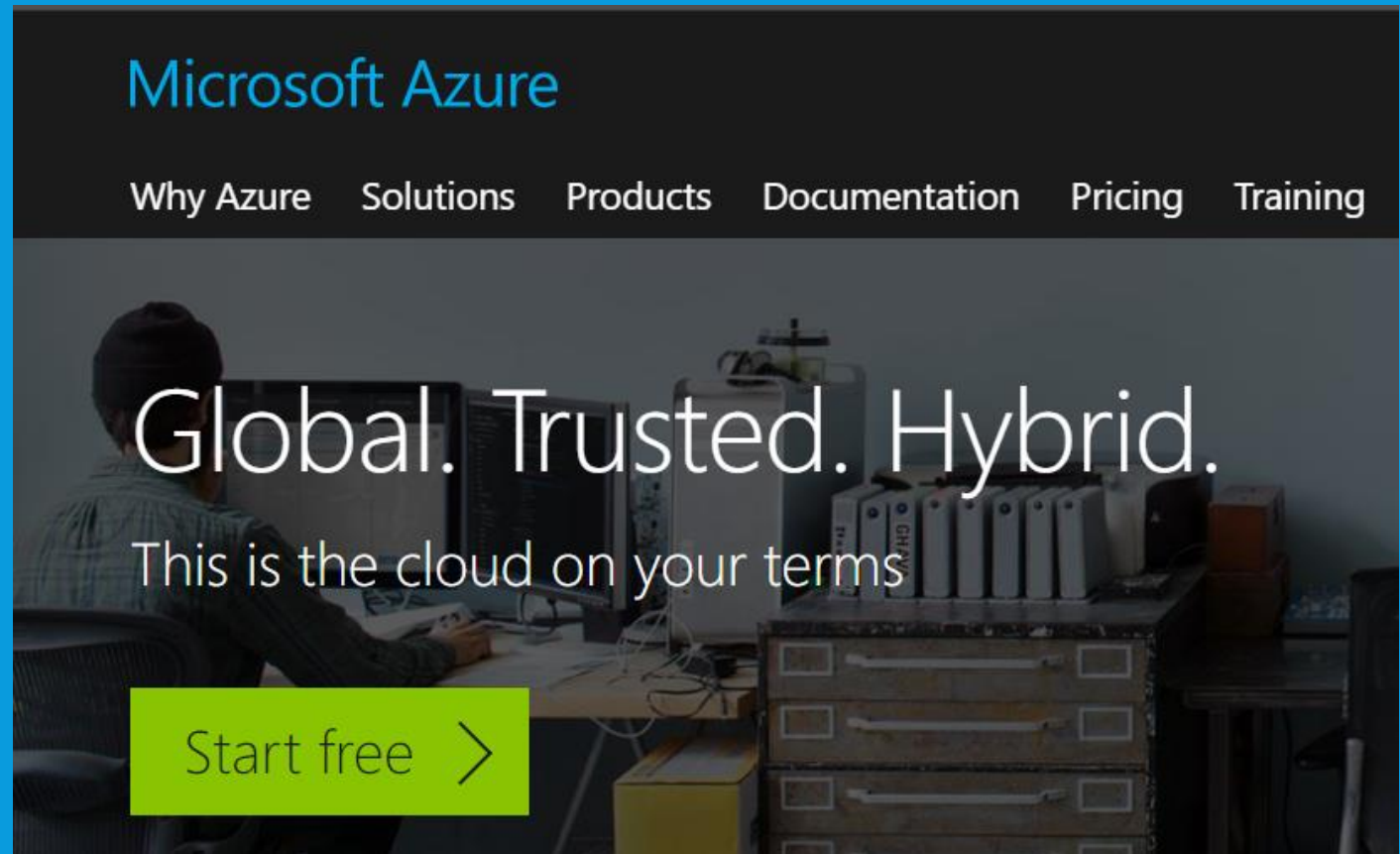


# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

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

Microsoft Azure

Search resources

Dashboard

+ New dashboard Edit dashboard Share Fullscreen Clone Delete

+ ←

All resources  
ALL SUBSCRIPTIONS

Refresh

- testrunning-vm1
- adev8602
- ddev2645
- dev
- dev3913
- dockerlab
- dockerlab\_OsDisk\_1\_6...
- dockerlab314

See more...

Service Health

Personalized guidance and support when issues in Azure services affect you. [Learn more](#)

Marketplace

Help + support

dockerlab

Stopped

Resources  
K8N

- dockerlab
- dockerlab\_OsDisk\_1\_6...
- dockerlab314

See more...

# 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

The screenshot displays the Microsoft Azure portal interface. On the left, the 'New' button is highlighted in the navigation pane. The main area shows the 'New' page with a search bar containing 'Ubuntu Server 16.04 LTS'. Below the search bar, the 'Azure Marketplace' section is visible, with 'Get started' highlighted. The 'Most recently created' section lists 'Ubuntu Server 16.04 LTS' with a 'Create' button. On the right, a detailed view of the 'Ubuntu Server 16.04 LTS' offering is shown, including the Canonical logo, a description of the VM, legal terms, and social media links. At the bottom of this panel, the 'Create' button is highlighted.

Microsoft Azure New > Ubuntu Server 16.04 LTS

Ubuntu Server 16.04 LTS  
Canonical

Ubuntu Server 16.04 LTS amd64 20170811 Public Azure, 20170811 Azure China, 20170811 Azure Germany, 20170619.1 Azure Gov. Ubuntu Server is the world's most popular Linux for cloud environments. Updates and patches for Ubuntu 16.04 will be available until April 2021. Ubuntu is the perfect virtual machine (VM) platform for all workloads from web applications to databases and Hadoop. For more information see [Ubuntu on Azure](#) and using Juju to create workloads.

Legal Terms

By clicking the Create button, I acknowledge that I am getting this software from Canonical and the legal terms of Canonical apply to it. Microsoft does not provide rights for third-party software. Also see the [privacy statement](#) from Canonical.

PUBLISHER Canonical

USEFUL LINKS [Documentation](#) [FAQ](#) [Pricing details](#)

Select a deployment model ●  
Resource Manager

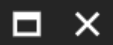
Create

# 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



## Create virtual machine



1

Basics  
Configure basic settings



2

Size  
Choose virtual machine size



3

Settings  
Configure optional features

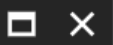


4

Purchase  
Ubuntu Server 16.04 LTS



## Basics



\* Name

hadoopdemo



VM disk type

SSD



\* User name

hadoopdemo



\* Authentication type

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

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 ★	D4S_V3 Standard ★	E2S_V3 Standard ★
2 vCPUs	4 vCPUs	2 vCPUs
8 GB	16 GB	16 GB
4 Data disks	8 Data disks	4 Data disks
4000 Max IOPS	8000 Max IOPS	4000 Max IOPS
16 GB Local SSD	32 GB Local SSD	32 GB Local SSD
Premium disk support	Premium disk support	Premium disk support
Load balancing	Load balancing	Load balancing
93.00 USD/MONTH (ESTIMATED)	186.00 USD/MONTH (ESTIMATED)	119.04 USD/MONTH (ESTIMATED)

Select





## 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

## 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](#)  
Standard D2s v3  
by Microsoft  
[Terms of use](#) | [privacy policy](#)

[Pricing details](#)

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)

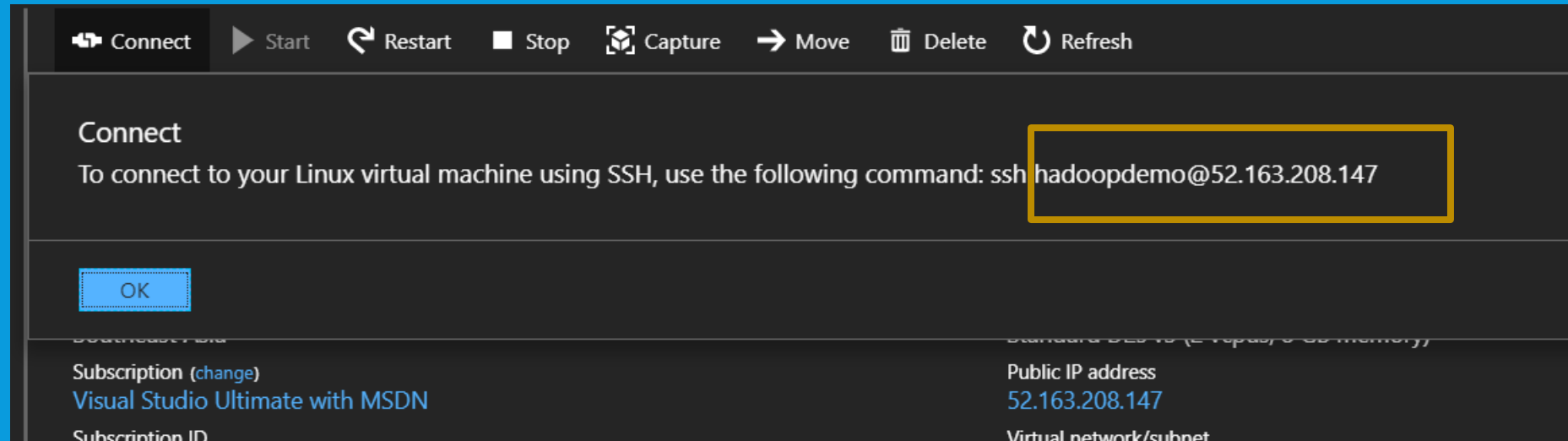
☐ I give Microsoft permission to use and share my contact information so that Microsoft or the Provider can contact me regarding this product and related products.

Purchase

[Download template and parameters](#)

# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- Get the IP address



Connect

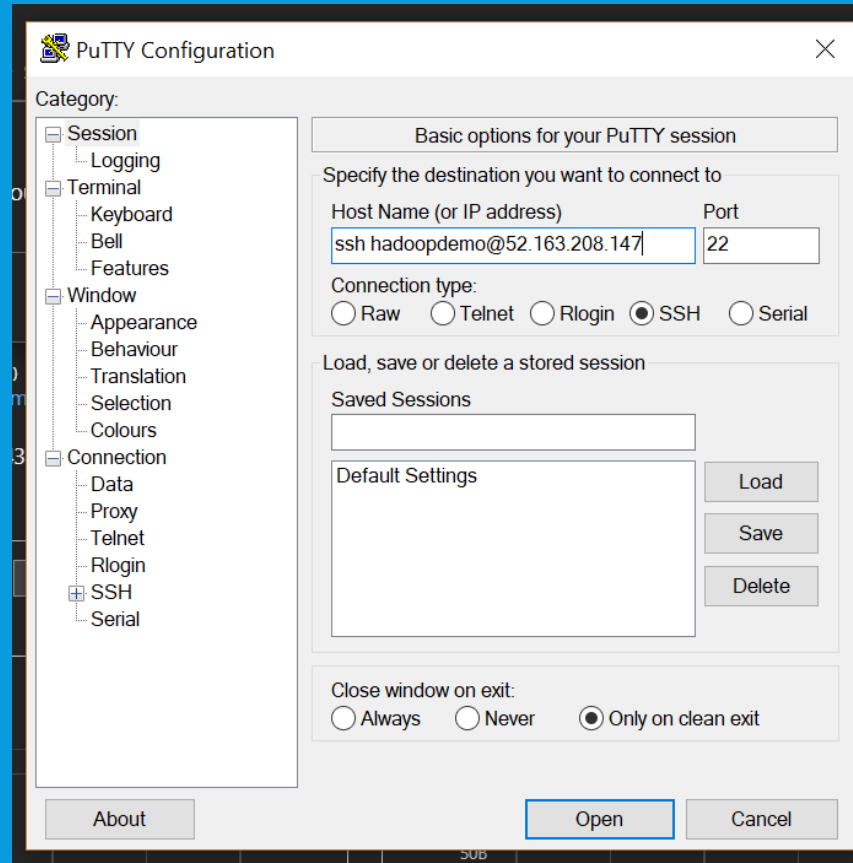
To connect to your Linux virtual machine using SSH, use the following command: `ssh hadoopdemo@52.163.208.147`

OK

Subscription (change)	Public IP address
Visual Studio Ultimate with MSDN	52.163.208.147
Subscription ID	Virtual network/subnet

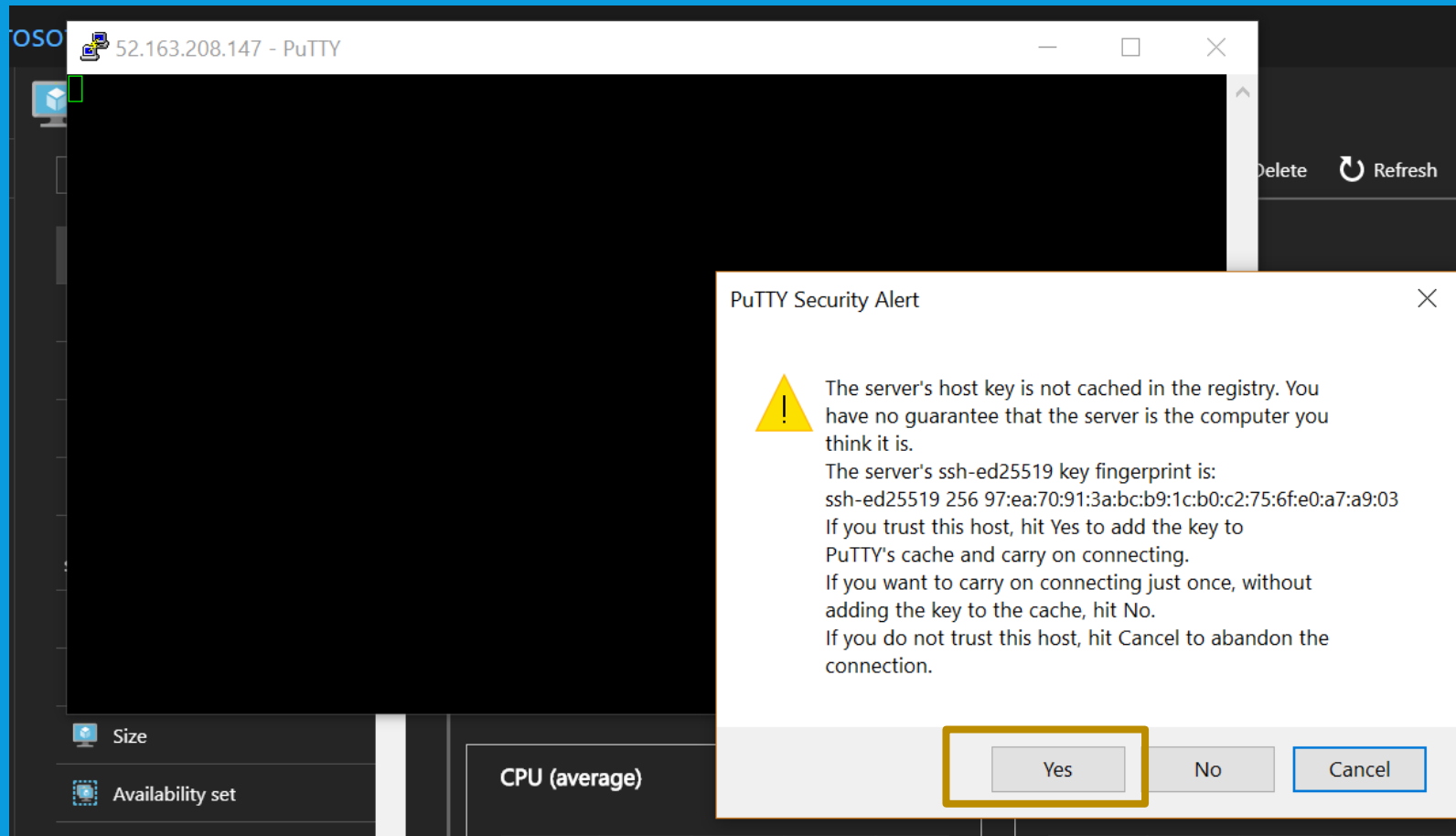
# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- Connect to an instance from Windows using Putty



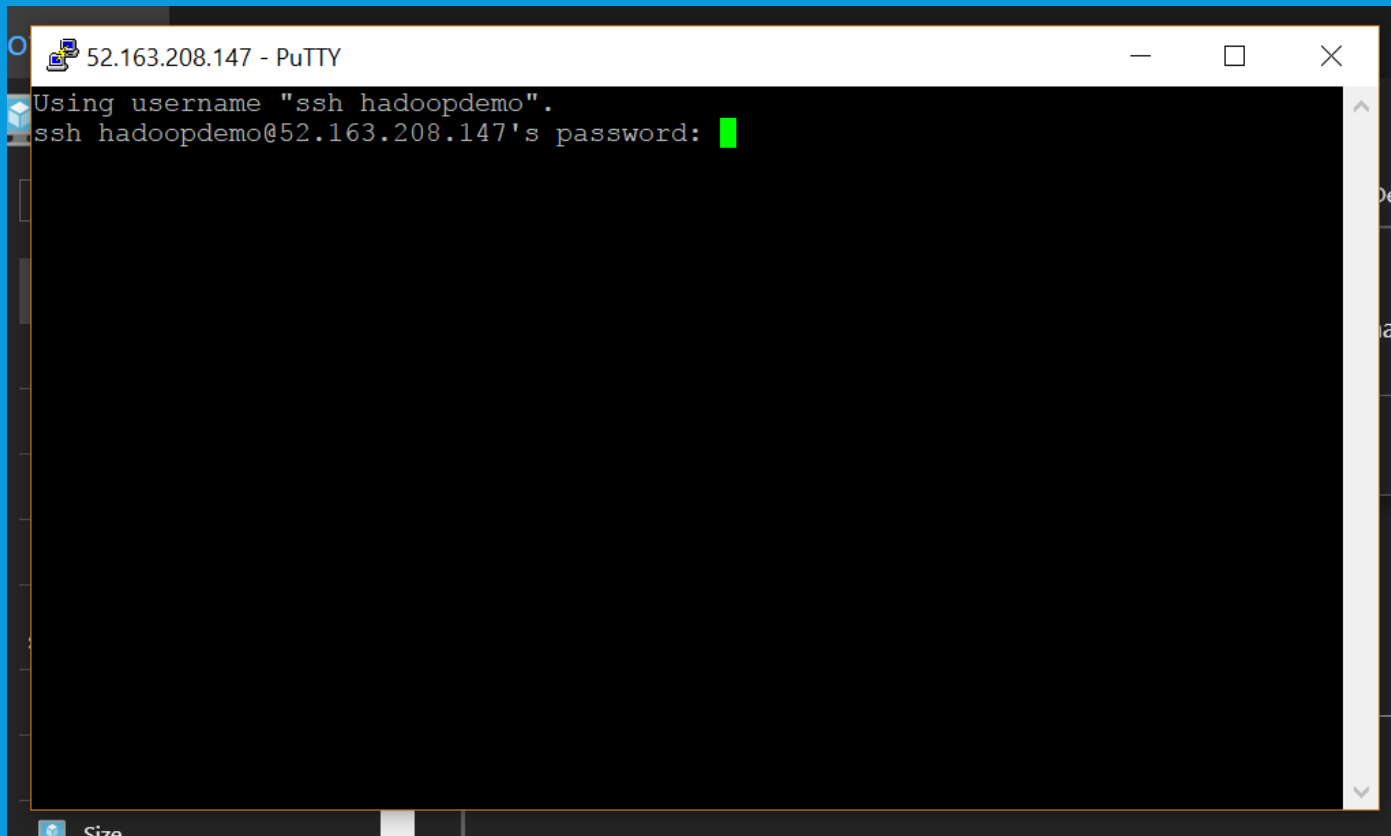
# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- Connect to an instance from Windows using Putty



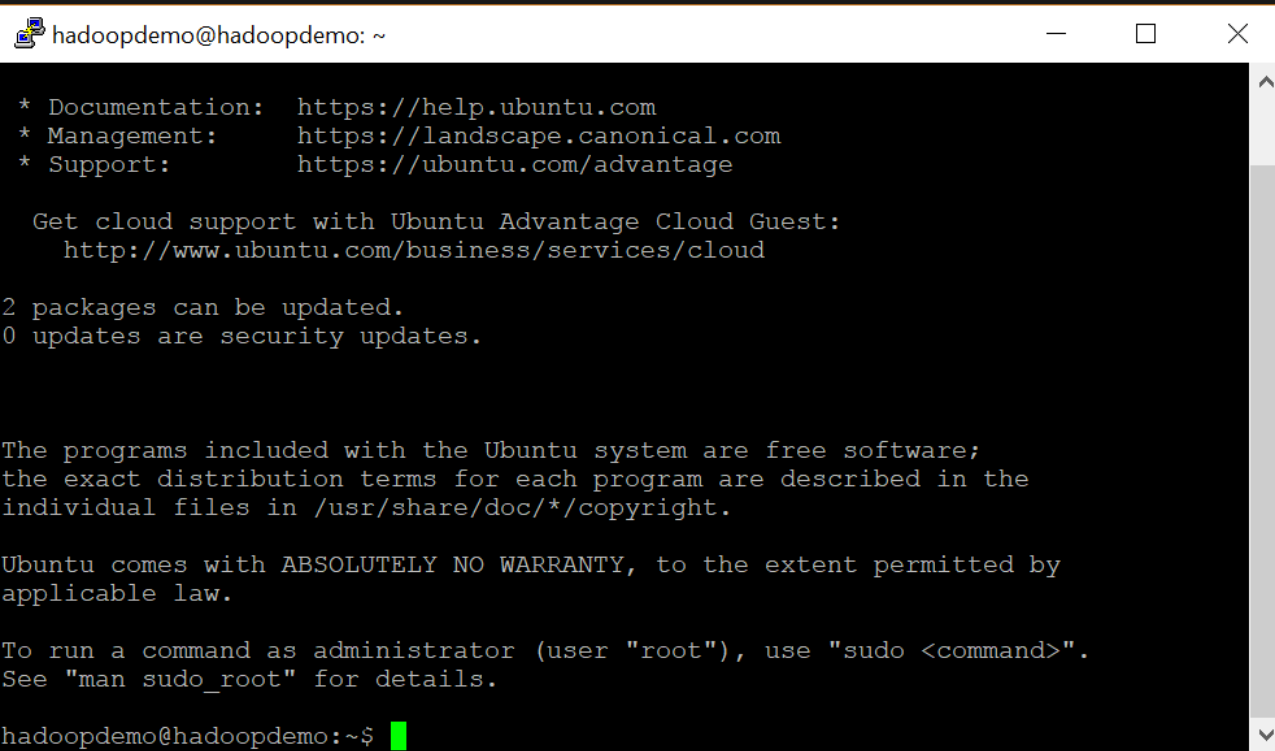
# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

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



# HANDS-ON LAUNCH A VIRTUAL MACHINE ON MICROSOFT AZURE

- Login

A terminal window titled 'hadoopdemo@hadoopdemo: ~' with standard window controls. The terminal displays Ubuntu system information, including documentation, management, and support links. It also shows a message about cloud support, package update status (2 packages can be updated, 0 security updates), and a disclaimer about warranty. The prompt 'hadoopdemo@hadoopdemo:~\$' is visible at the bottom with a green cursor.

```
hadoopdemo@hadoopdemo: ~  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:       https://ubuntu.com/advantage  
  
Get cloud support with Ubuntu Advantage Cloud Guest:  
http://www.ubuntu.com/business/services/cloud  
  
2 packages can be updated.  
0 updates are security updates.  
  
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  
applicable law.  
  
To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo_root" for details.  
  
hadoopdemo@hadoopdemo:~$
```

# HANDS-ON INSTALLING HADOOP

## Installing Hadoop

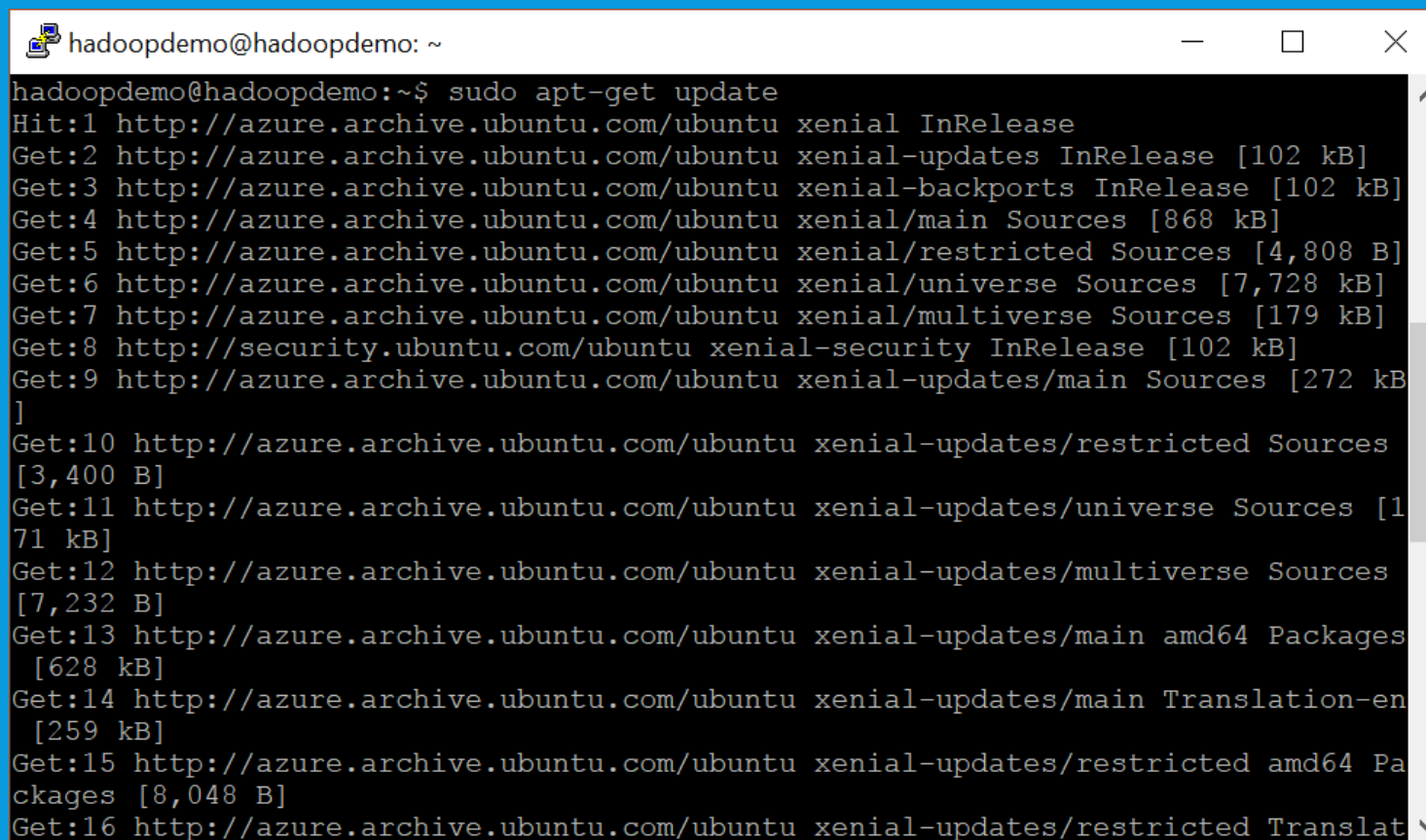
1. Update System Software Repository
2. Configuring SSH
3. Installing Java
4. Download/Extract Hadoop
5. Installing Hadoop
6. Configure Hadoop
7. Formatting Namenode
8. Starting Hadoop
9. Accessing Hadoop Web Console
10. Stopping Hadoop



# HANDS-ON INSTALLING HADOOP

## 1. Update System Software Repository

- `$ sudo apt-get update`

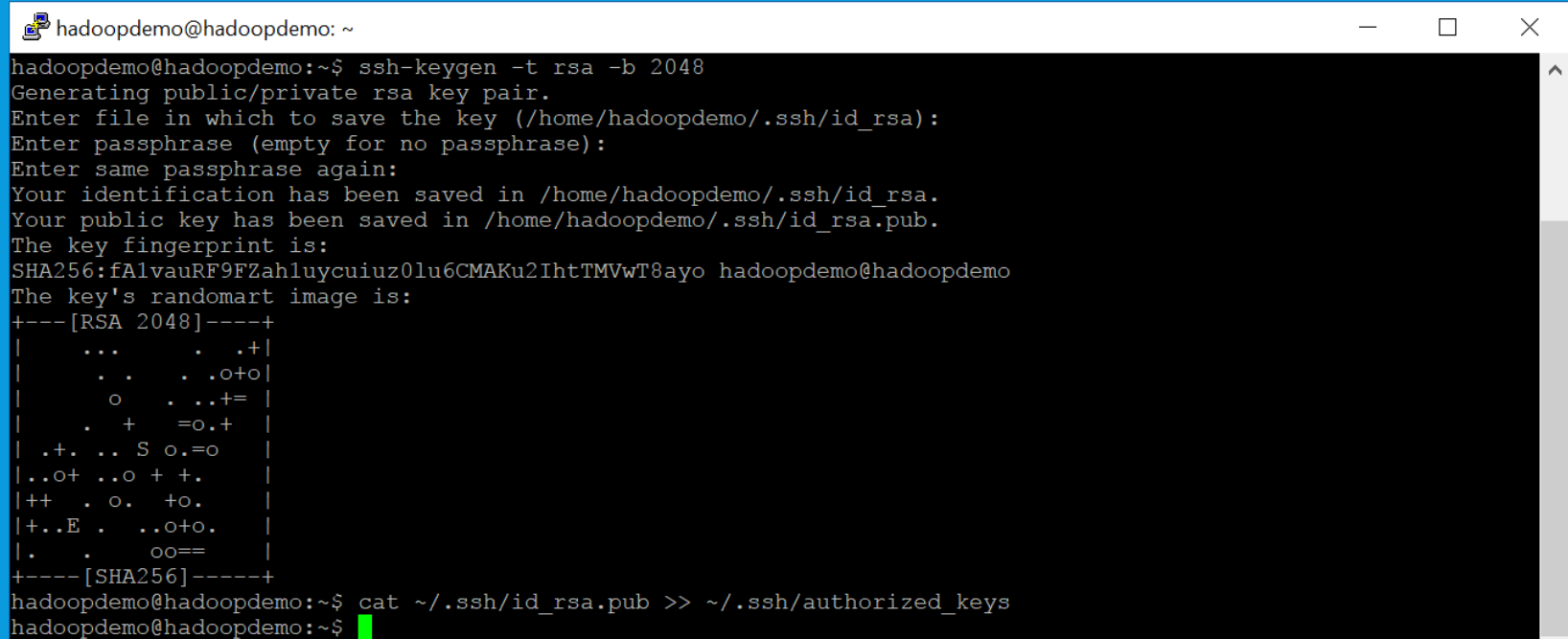
A terminal window titled 'hadoopdemo@hadoopdemo: ~' with standard window controls. The terminal displays the output of the command 'sudo apt-get update'. The output shows a series of 'Hit' and 'Get' messages for various Ubuntu repositories, including xenial InRelease, xenial-updates InRelease, xenial-backports InRelease, xenial/main Sources, xenial/restricted Sources, xenial/universe Sources, xenial/multiverse Sources, xenial-security InRelease, and xenial-updates/main Sources. It continues with xenial-updates/restricted Sources, xenial-updates/universe Sources, xenial-updates/multiverse Sources, xenial-updates/main amd64 Packages, xenial-updates/main Translation-en, xenial-updates/restricted amd64 Packages, and xenial-updates/restricted Translat. The window has a scrollbar on the right side.

```
hadoopdemo@hadoopdemo:~$ sudo apt-get update
Hit:1 http://azure.archive.ubuntu.com/ubuntu xenial InRelease
Get:2 http://azure.archive.ubuntu.com/ubuntu xenial-updates InRelease [102 kB]
Get:3 http://azure.archive.ubuntu.com/ubuntu xenial-backports InRelease [102 kB]
Get:4 http://azure.archive.ubuntu.com/ubuntu xenial/main Sources [868 kB]
Get:5 http://azure.archive.ubuntu.com/ubuntu xenial/restricted Sources [4,808 B]
Get:6 http://azure.archive.ubuntu.com/ubuntu xenial/universe Sources [7,728 kB]
Get:7 http://azure.archive.ubuntu.com/ubuntu xenial/multiverse Sources [179 kB]
Get:8 http://security.ubuntu.com/ubuntu xenial-security InRelease [102 kB]
Get:9 http://azure.archive.ubuntu.com/ubuntu xenial-updates/main Sources [272 kB]
]
Get:10 http://azure.archive.ubuntu.com/ubuntu xenial-updates/restricted Sources
[3,400 B]
Get:11 http://azure.archive.ubuntu.com/ubuntu xenial-updates/universe Sources [1
71 kB]
Get:12 http://azure.archive.ubuntu.com/ubuntu xenial-updates/multiverse Sources
[7,232 B]
Get:13 http://azure.archive.ubuntu.com/ubuntu xenial-updates/main amd64 Packages
[628 kB]
Get:14 http://azure.archive.ubuntu.com/ubuntu xenial-updates/main Translation-en
[259 kB]
Get:15 http://azure.archive.ubuntu.com/ubuntu xenial-updates/restricted amd64 Pa
ckages [8,048 B]
Get:16 http://azure.archive.ubuntu.com/ubuntu xenial-updates/restricted Translat
```

# HANDS-ON INSTALLING HADOOP

## 2. Configuring SSH

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

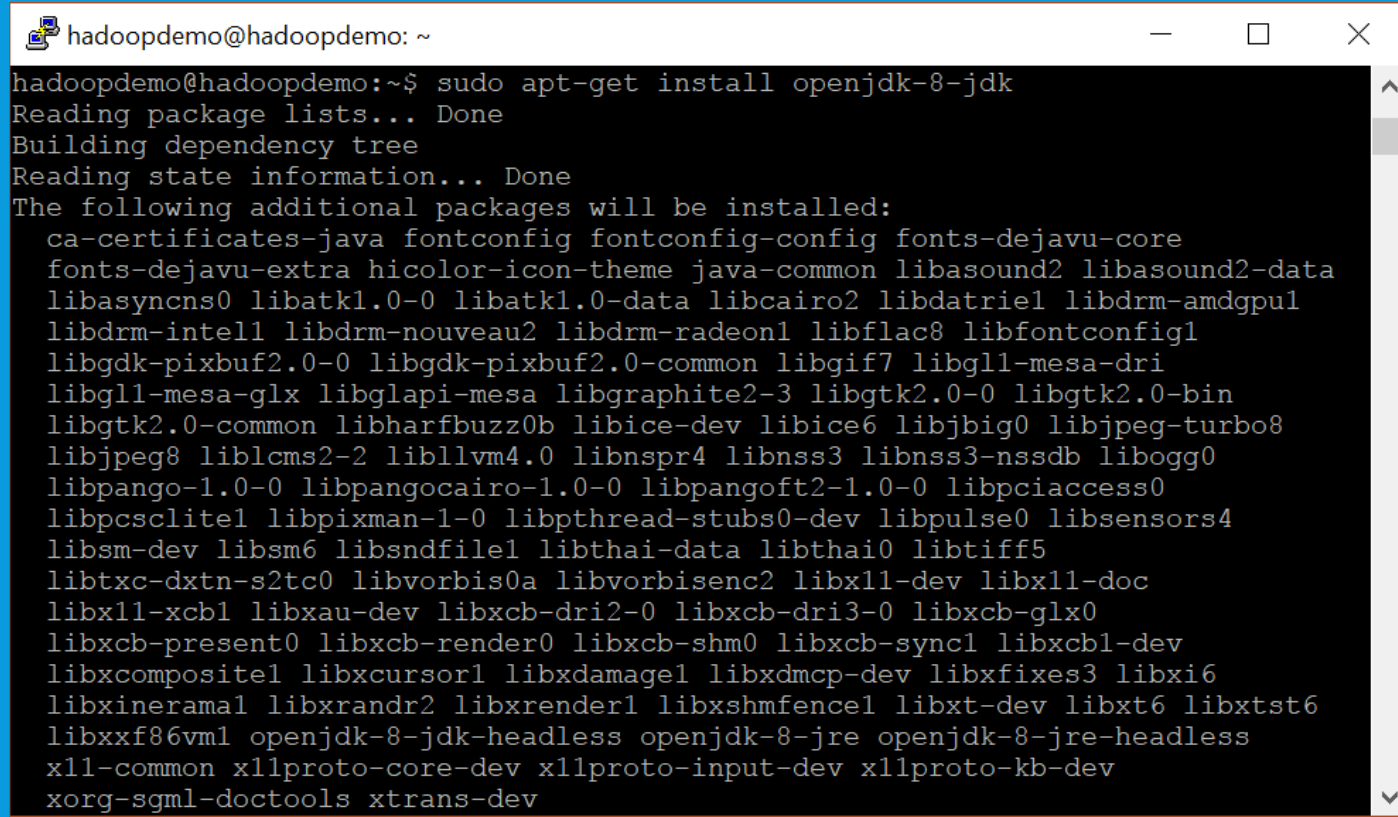


```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ ssh-keygen -t rsa -b 2048  
Generating public/private rsa key pair.  
Enter file in which to save the key (/home/hadoopdemo/.ssh/id_rsa):  
Enter passphrase (empty for no passphrase):  
Enter same passphrase again:  
Your identification has been saved in /home/hadoopdemo/.ssh/id_rsa.  
Your public key has been saved in /home/hadoopdemo/.ssh/id_rsa.pub.  
The key fingerprint is:  
SHA256:fAlvauRF9FZahluycuiuz0lu6CMAKu2IhtTMVwT8ayo hadoopdemo@hadoopdemo  
The key's randomart image is:  
+---[RSA 2048]---+  
|    . . . . .+ |  
|    . . . . .o+o |  
|    o . . . += |  
|    . + =o.+ |  
| .+. . . S o.=o |  
|..o+ ..o + +. |  
|++ . o. +o. |  
|+..E . ..o+o. |  
|. . . oo== |  
+----[SHA256]-----+  
hadoopdemo@hadoopdemo:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
hadoopdemo@hadoopdemo:~$
```

# HANDS-ON INSTALLING HADOOP

## 3. Installing Java

```
$ sudo apt-get install openjdk-8-jdk
```



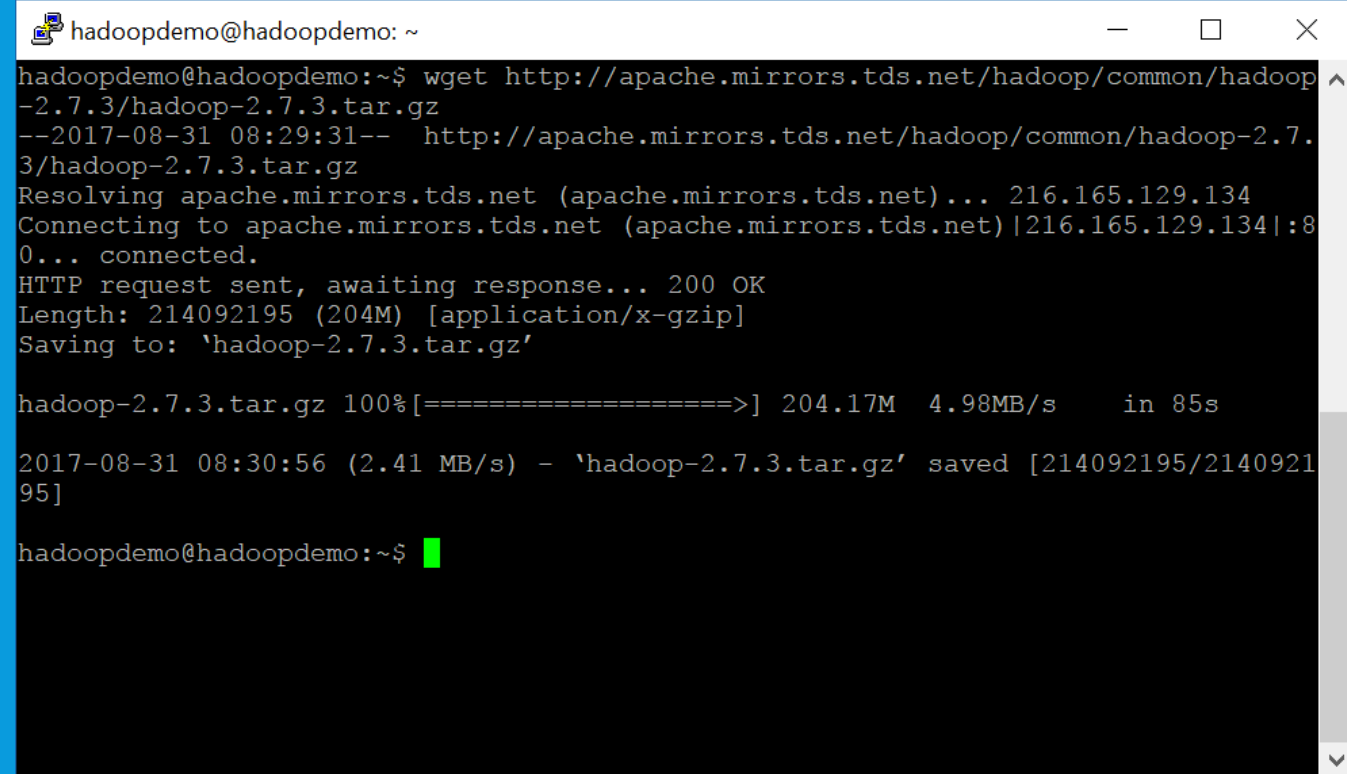
```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ sudo apt-get install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  ca-certificates-java fontconfig fontconfig-config fonts-dejavu-core  
  fonts-dejavu-extra hicolor-icon-theme java-common libasound2 libasound2-data  
  libasyncns0 libatk1.0-0 libatk1.0-data libcairo2 libdatriel libdrm-amdgpu1  
  libdrm-intel1 libdrm-nouveau2 libdrm-radeon1 libflac8 libfontconfig1  
  libgdk-pixbuf2.0-0 libgdk-pixbuf2.0-common libgif7 libgl1-mesa-dri  
  libgl1-mesa-glx libglapi-mesa libgraphite2-3 libgtk2.0-0 libgtk2.0-bin  
  libgtk2.0-common libharfbuzz0b libice-dev libice6 libjbig0 libjpeg-turbo8  
  libjpeg8 liblcms2-2 liblvm4.0 libnspr4 libnss3 libnss3-nssdb libogg0  
  libpango-1.0-0 libpangocairo-1.0-0 libpangoft2-1.0-0 libpciaccess0  
  libpcsclite1 libpixmap-1-0 libpthread-stubs0-dev libpulse0 libsensors4  
  libsm-dev libsm6 libsndfile1 libthai-data libthai0 libtiff5  
  libtxc-dxtn-s2tc0 libvorbis0a libvorbisenc2 libx11-dev libx11-doc  
  libx11-xcb1 libxau-dev libxcb-dri2-0 libxcb-dri3-0 libxcb-glx0  
  libxcb-present0 libxcb-render0 libxcb-shm0 libxcb-sync1 libxcb1-dev  
  libxcomposit1 libxcursor1 libxdamage1 libxdmcp-dev libxfixes3 libxi6  
  libxinerama1 libxrandr2 libxrender1 libxshmfence1 libxt-dev libxt6 libxtst6  
  libxxf86vm1 openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
  x11-common x11proto-core-dev x11proto-input-dev x11proto-kb-dev  
  xorg-sgml-doctools xtrans-dev
```

# 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
```



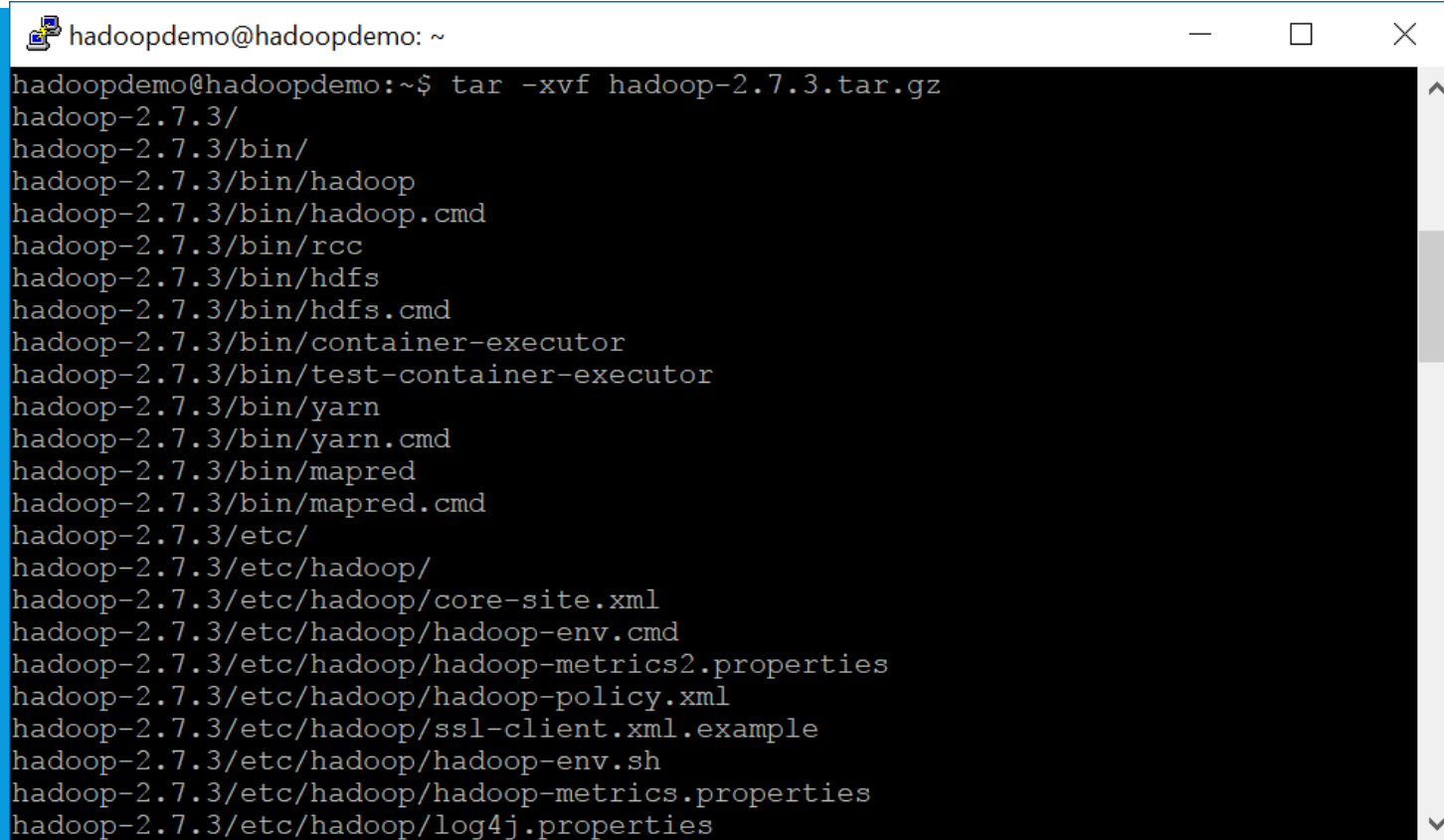
```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ wget http://apache.mirrors.tds.net/hadoop/common/hadoop-  
2.7.3/hadoop-2.7.3.tar.gz  
--2017-08-31 08:29:31-- http://apache.mirrors.tds.net/hadoop/common/hadoop-2.7.  
3/hadoop-2.7.3.tar.gz  
Resolving apache.mirrors.tds.net (apache.mirrors.tds.net)... 216.165.129.134  
Connecting to apache.mirrors.tds.net (apache.mirrors.tds.net)|216.165.129.134|:8  
0... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 214092195 (204M) [application/x-gzip]  
Saving to: 'hadoop-2.7.3.tar.gz'  
  
hadoop-2.7.3.tar.gz 100%[=====>] 204.17M 4.98MB/s in 85s  
  
2017-08-31 08:30:56 (2.41 MB/s) - 'hadoop-2.7.3.tar.gz' saved [214092195/2140921  
95]  
  
hadoopdemo@hadoopdemo:~$
```

# HANDS-ON INSTALLING HADOOP

- **4. Download/Extract Hadoop**

- Extract Hadoop

```
$ tar -xvf hadoop-2.7.3.tar.gz
```



```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ tar -xvf hadoop-2.7.3.tar.gz  
hadoop-2.7.3/  
hadoop-2.7.3/bin/  
hadoop-2.7.3/bin/hadoop  
hadoop-2.7.3/bin/hadoop.cmd  
hadoop-2.7.3/bin/rcc  
hadoop-2.7.3/bin/hdfs  
hadoop-2.7.3/bin/hdfs.cmd  
hadoop-2.7.3/bin/container-executor  
hadoop-2.7.3/bin/test-container-executor  
hadoop-2.7.3/bin/yarn  
hadoop-2.7.3/bin/yarn.cmd  
hadoop-2.7.3/bin/mapred  
hadoop-2.7.3/bin/mapred.cmd  
hadoop-2.7.3/etc/  
hadoop-2.7.3/etc/hadoop/  
hadoop-2.7.3/etc/hadoop/core-site.xml  
hadoop-2.7.3/etc/hadoop/hadoop-env.cmd  
hadoop-2.7.3/etc/hadoop/hadoop-metrics2.properties  
hadoop-2.7.3/etc/hadoop/hadoop-policy.xml  
hadoop-2.7.3/etc/hadoop/ssl-client.xml.example  
hadoop-2.7.3/etc/hadoop/hadoop-env.sh  
hadoop-2.7.3/etc/hadoop/hadoop-metrics.properties  
hadoop-2.7.3/etc/hadoop/log4j.properties
```

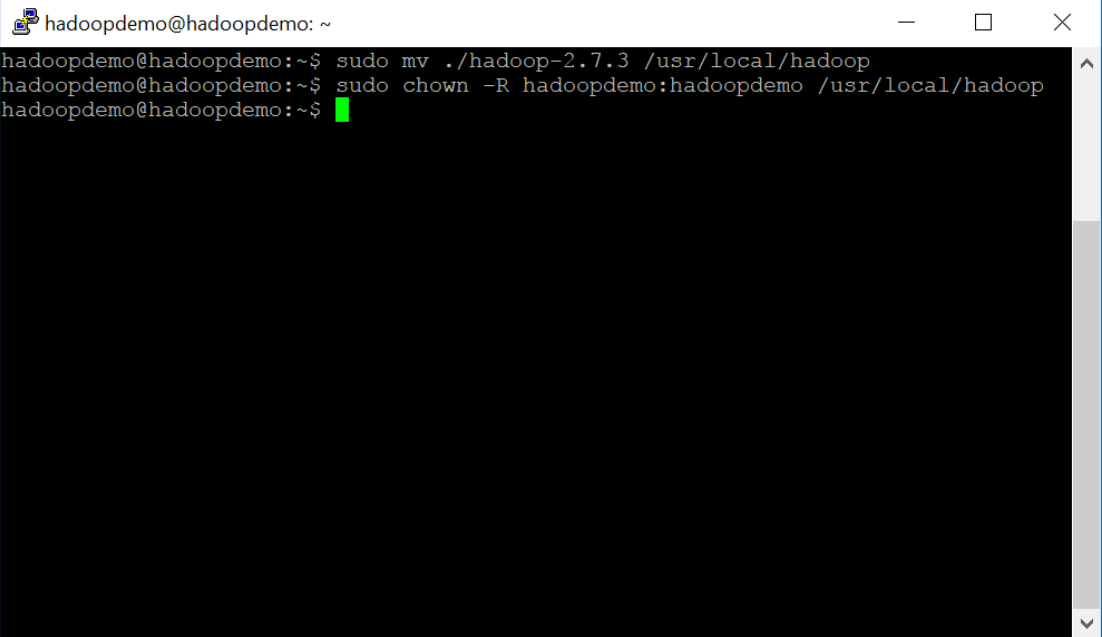
# HANDS-ON INSTALLING HADOOP

## 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
```



```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ sudo mv ./hadoop-2.7.3 /usr/local/hadoop  
hadoopdemo@hadoopdemo:~$ sudo chown -R hadoopdemo:hadoopdemo /usr/local/hadoop  
hadoopdemo@hadoopdemo:~$
```

# HANDS-ON INSTALLING HADOOP

## 5. Installing Hadoop

- Set startup path and configuration bash shell

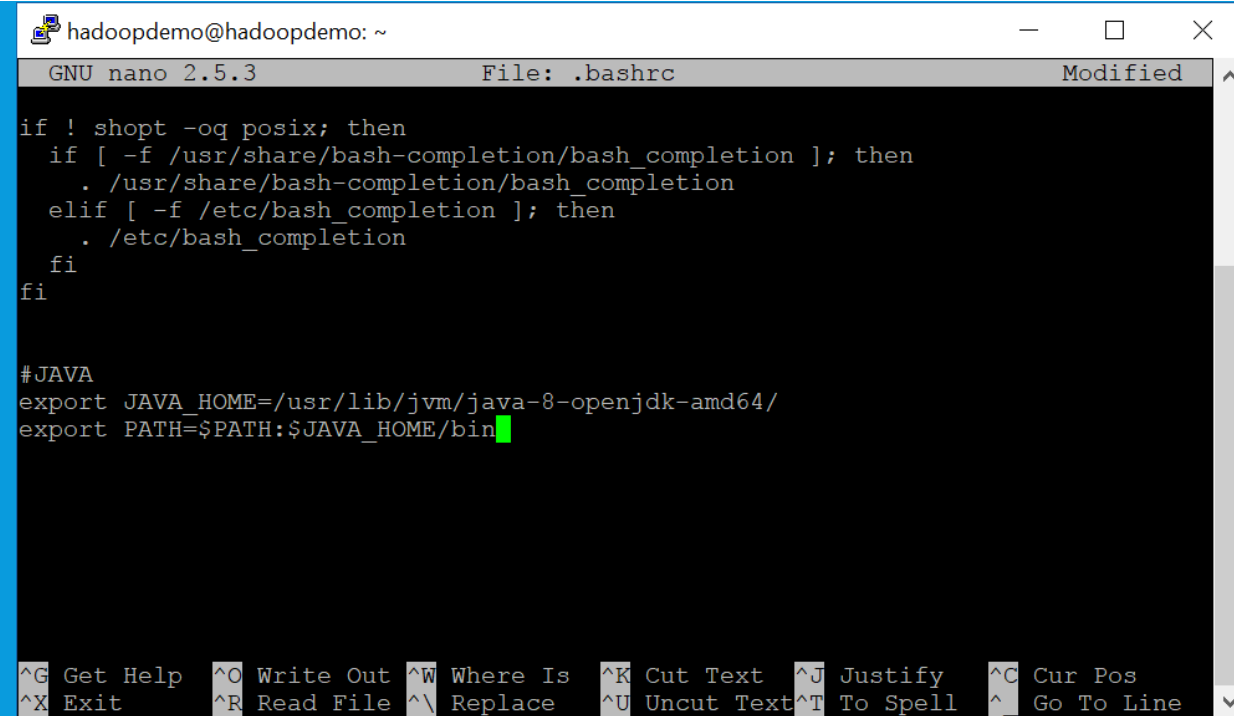
```
$ sudo nano .bashrc
```

- Configuration java path

```
#JAVA
```

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

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

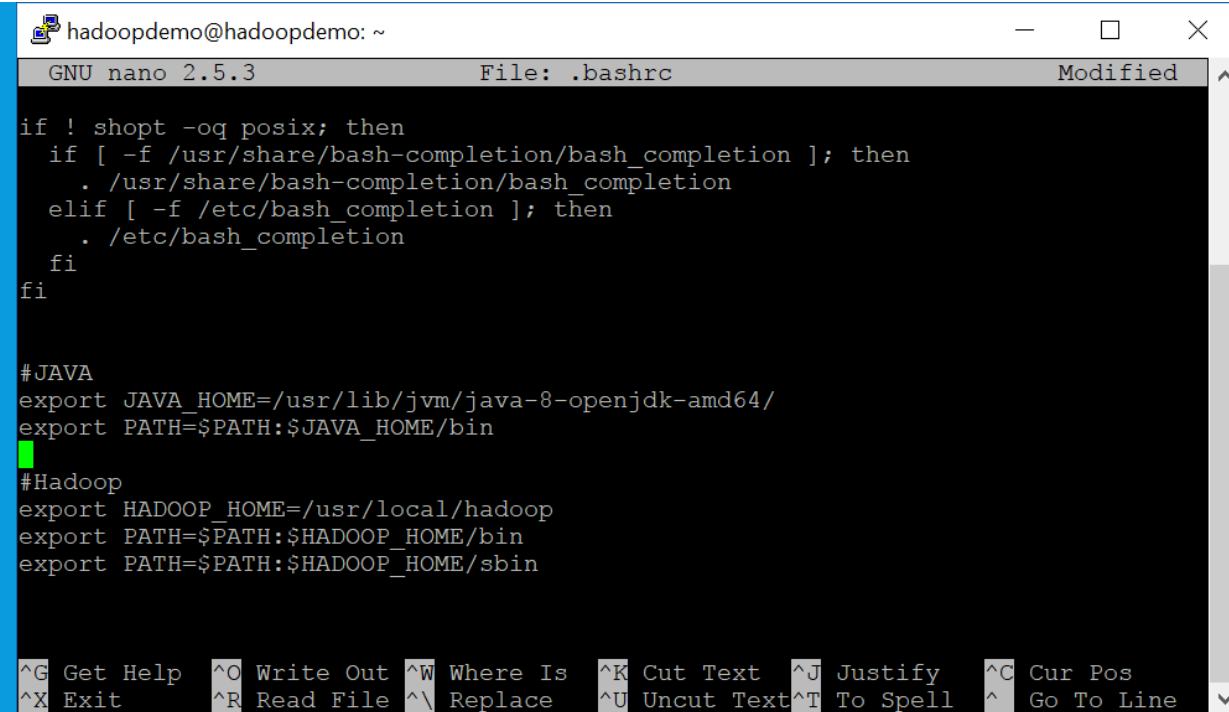


```
hadoopdemo@hadoopdemo: ~  
GNU nano 2.5.3 File: .bashrc Modified  
  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
#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
```



The screenshot shows a terminal window titled 'hadoopdemo@hadoopdemo: ~'. The editor is GNU nano 2.5.3, editing the file '.bashrc'. The content of the file is as follows:

```
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

#JAVA
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
export PATH=$PATH:$JAVA_HOME/bin

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

The bottom of the window shows a status bar with various keyboard shortcuts: ^G Get Help, ^O Write Out, ^W Where Is, ^K Cut Text, ^J Justify, ^C Cur Pos, ^X Exit, ^R Read File, ^\ Replace, ^U Uncut Text, ^T To Spell, and ^\_ Go To Line.

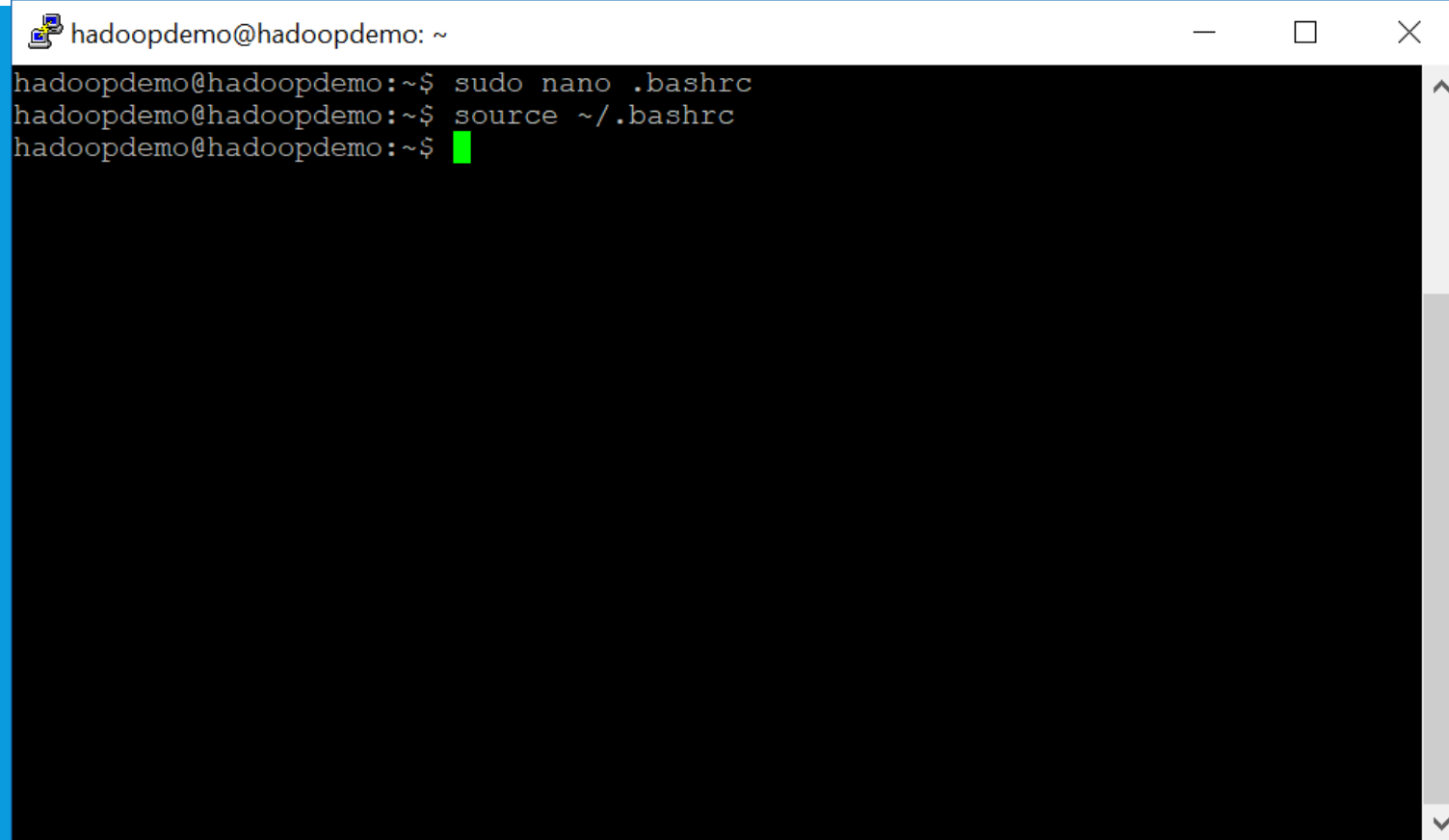


# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Execute environment variables:

*\$ source ~/.bashrc*



```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ sudo nano .bashrc  
hadoopdemo@hadoopdemo:~$ source ~/.bashrc  
hadoopdemo@hadoopdemo:~$
```

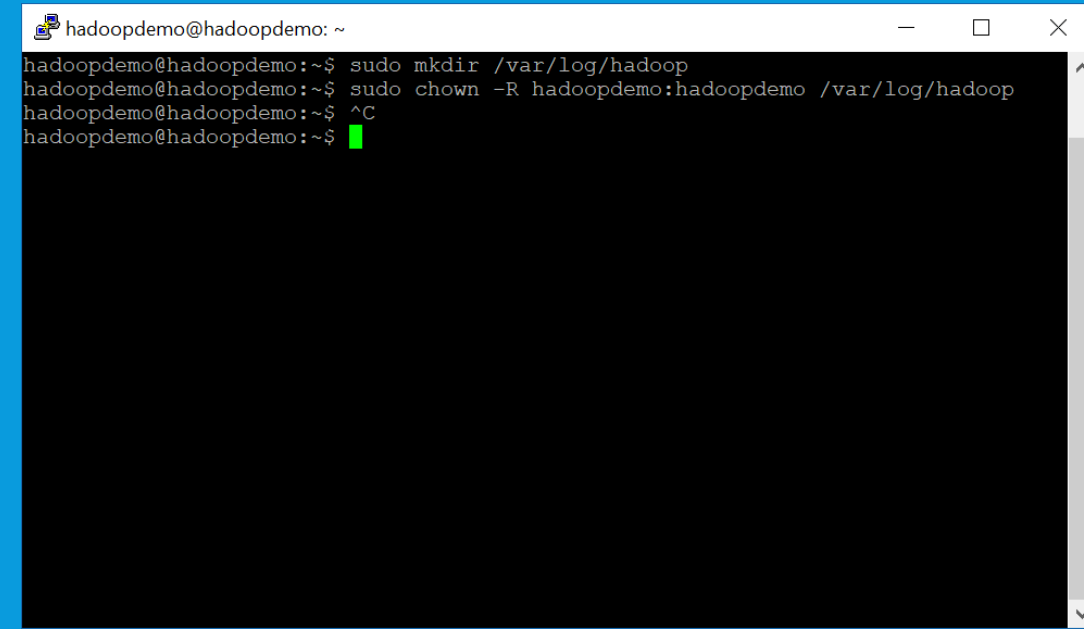
# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Create hadoop log path

```
$ sudo mkdir /var/log/hadoop
```

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



```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ sudo mkdir /var/log/hadoop  
hadoopdemo@hadoopdemo:~$ sudo chown -R hadoopdemo:hadoopdemo /var/log/hadoop  
hadoopdemo@hadoopdemo:~$ ^C  
hadoopdemo@hadoopdemo:~$
```

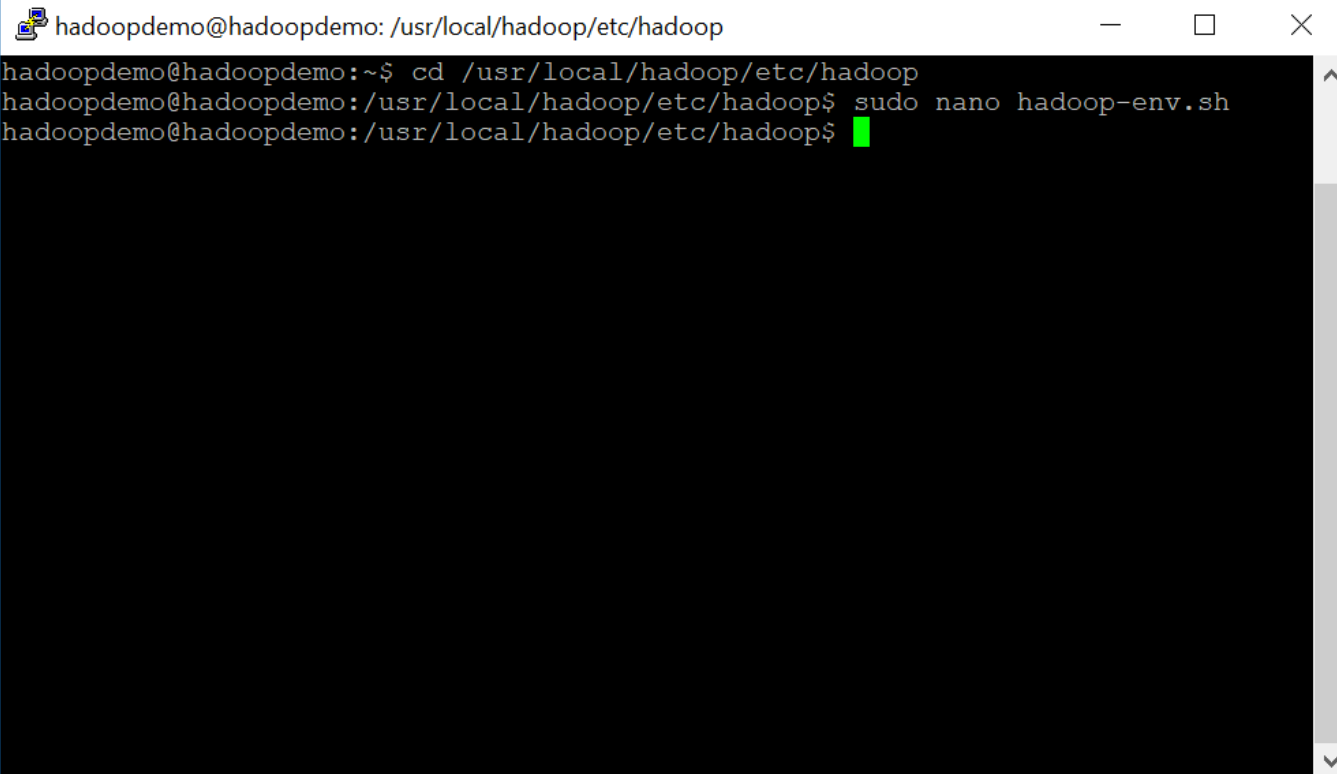
# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Edit hadoop shell script

```
$ cd /usr/local/hadoop/etc/hadoop
```

```
$ sudo nano hadoop-env.sh
```

A terminal window with a title bar showing 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The terminal has a black background with white text. It shows the user navigating to the Hadoop configuration directory and opening the 'hadoop-env.sh' file with 'nano' as root. A green cursor is visible at the end of the last command line.

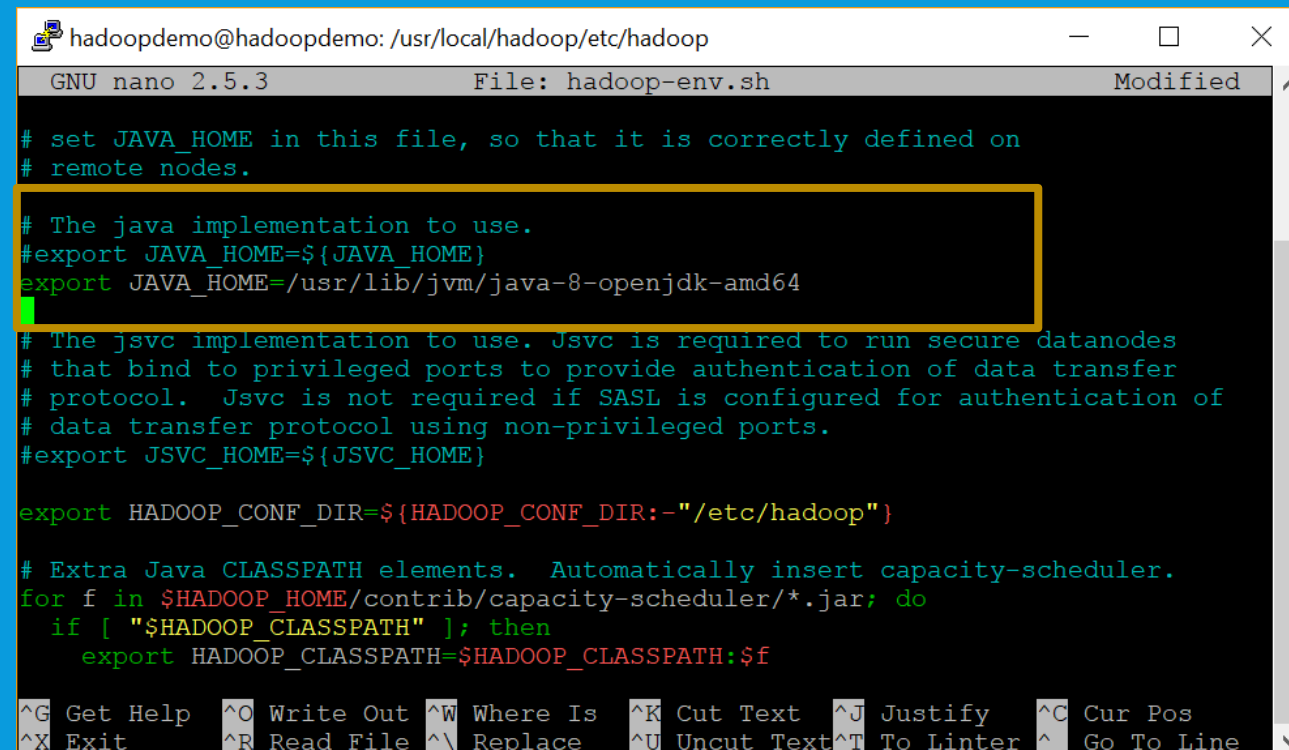
```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:~$ cd /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo nano hadoop-env.sh
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$
```

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Configuration JAVA\_HOME

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



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
GNU nano 2.5.3 File: hadoop-env.sh Modified
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol. Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=${JSVC_HOME}

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}

# Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
for f in $HADOOP_HOME/contrib/capacity-scheduler/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  fi
done

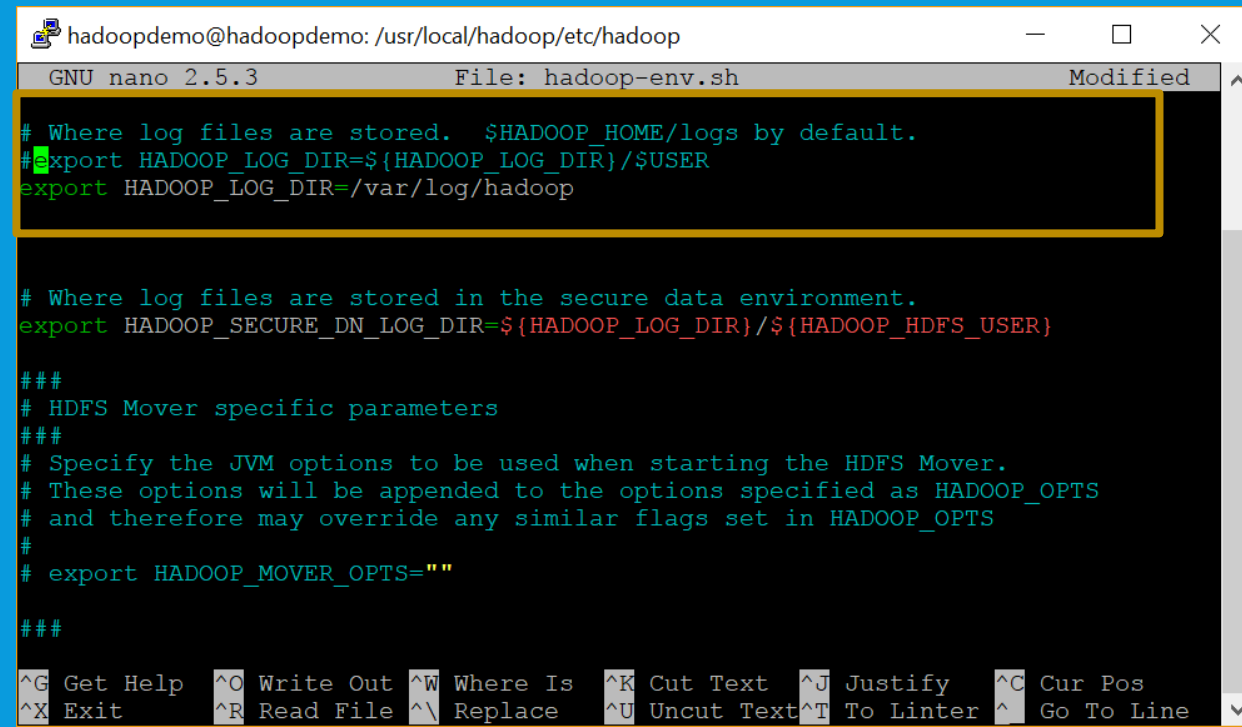
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^_ Go To Line
```

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Configuration hadoop log

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



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
GNU nano 2.5.3 File: hadoop-env.sh Modified
# Where log files are stored. $HADOOP_HOME/logs by default.
# export HADOOP_LOG_DIR=${HADOOP_LOG_DIR}/${USER}
export HADOOP_LOG_DIR=/var/log/hadoop

# Where log files are stored in the secure data environment.
export HADOOP_SECURE_DN_LOG_DIR=${HADOOP_LOG_DIR}/${HADOOP_HDFS_USER}

###
# HDFS Mover specific parameters
###
# Specify the JVM options to be used when starting the HDFS Mover.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# export HADOOP_MOVER_OPTS=""

###
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^_ Go To Line
```

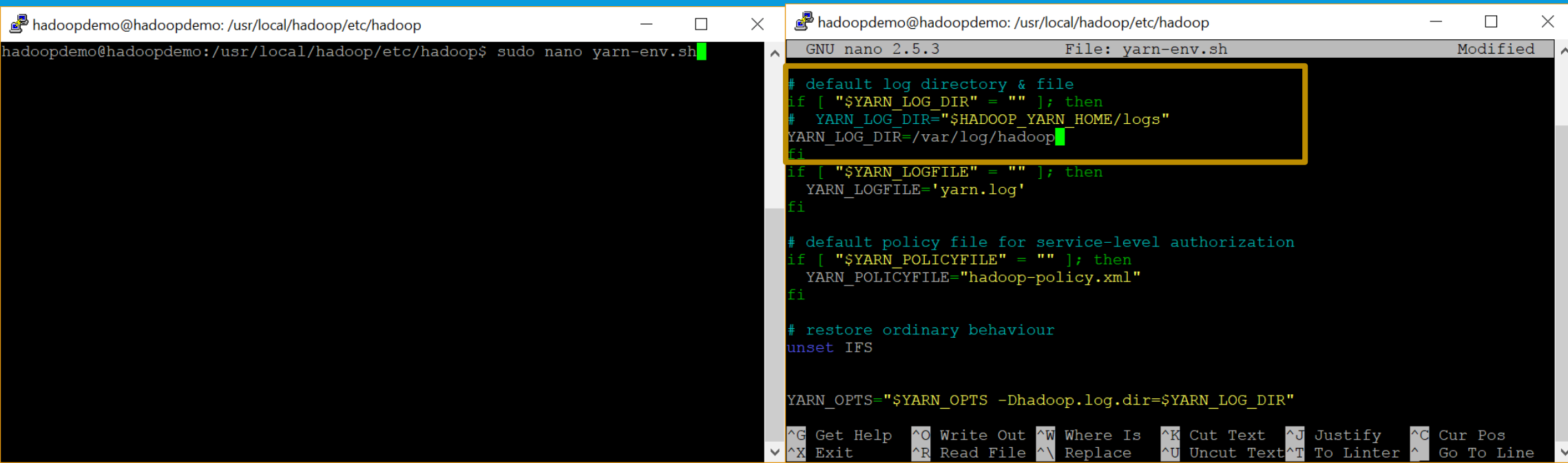
# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Edit Yarn shell script – log location to another directory

```
$ sudo nano yarn-env.sh
```

```
YARN_LOG_DIR=/var/log/hadoop
```



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo nano yarn-env.sh
```

```
GNU nano 2.5.3 File: yarn-env.sh Modified
# default log directory & file
if [ "$YARN_LOG_DIR" = "" ]; then
#   YARN_LOG_DIR="$HADOOP_YARN_HOME/logs"
YARN_LOG_DIR=/var/log/hadoop
fi
if [ "$YARN_LOGFILE" = "" ]; then
  YARN_LOGFILE='yarn.log'
fi

# default policy file for service-level authorization
if [ "$YARN_POLICYFILE" = "" ]; then
  YARN_POLICYFILE="hadoop-policy.xml"
fi

# restore ordinary behaviour
unset IFS

YARN_OPTS="$YARN_OPTS -Dhadoop.log.dir=$YARN_LOG_DIR"
```

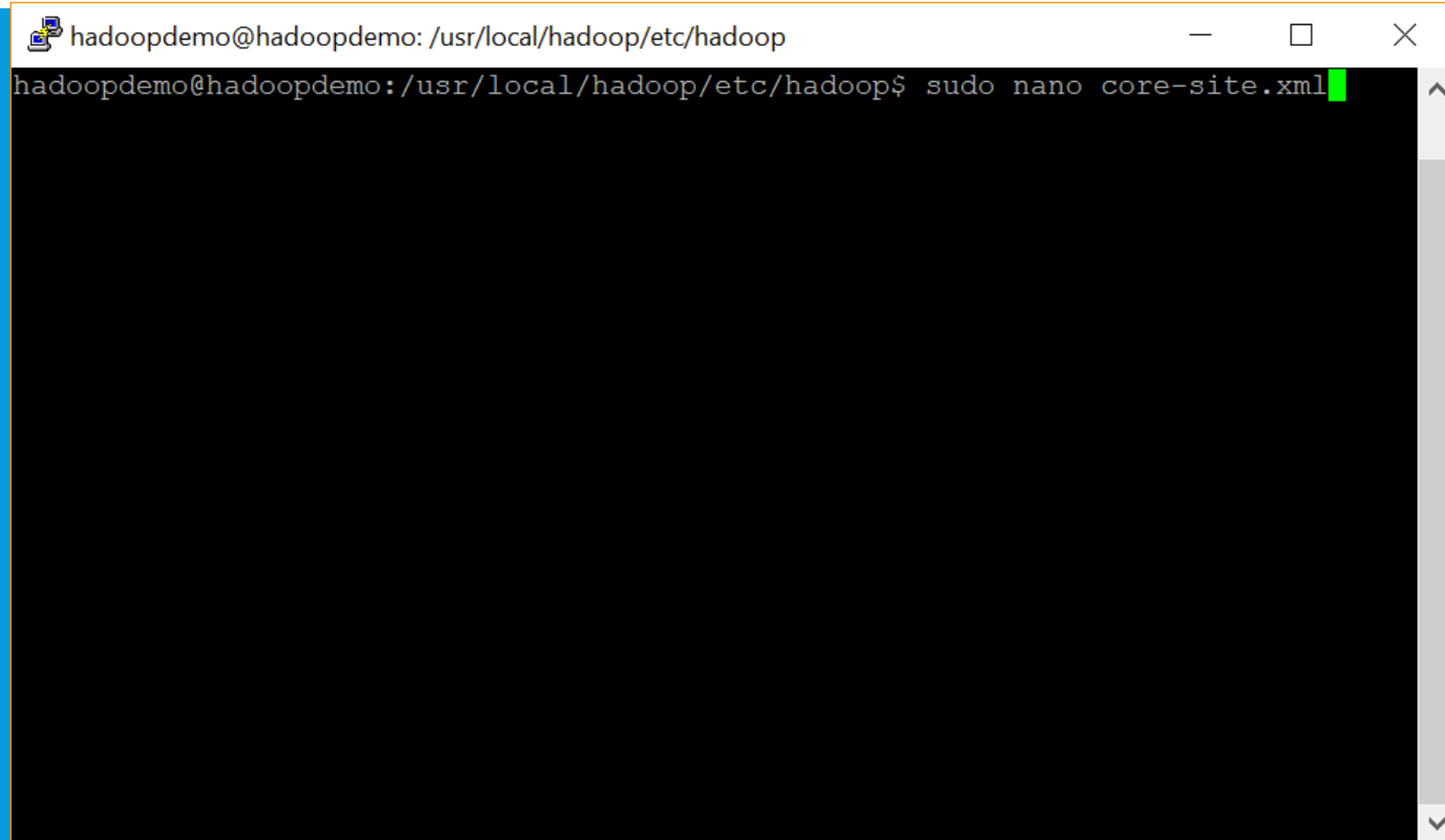
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Linter ^\_ Go To Line

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Edit hadoop - core-site.xml

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

A terminal window with a white title bar and standard window controls. The title bar text is 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The terminal content shows the command 'sudo nano core-site.xml' being entered at the prompt. A green cursor is visible at the end of the command line.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo nano core-site.xml
```

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- My Ip is 10.0.17.4

```
<configuration>
```

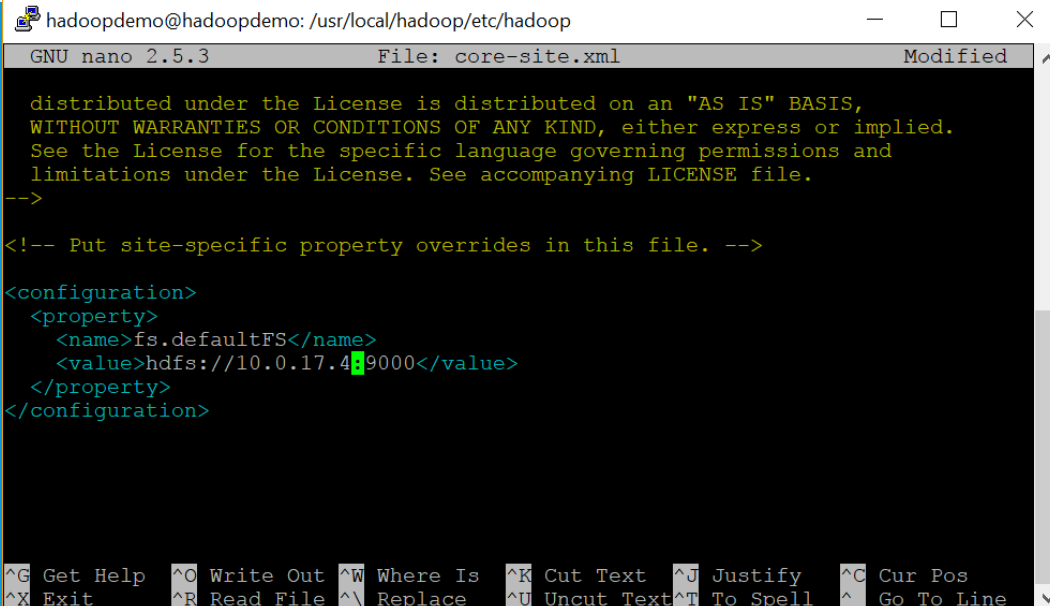
```
  <property>
```

```
    <name>fs.defaultFS</name>
```

```
    <value>hdfs://10.0.17.4:9000</value>
```

```
  </property>
```

```
</configuration>
```



The screenshot shows a terminal window with the title 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The editor is GNU nano 2.5.3, editing the file 'core-site.xml'. The content of the file is as follows:

```
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
  <property>  
    <name>fs.defaultFS</name>  
    <value>hdfs://10.0.17.4:9000</value>  
  </property>  
</configuration>
```

The bottom of the window shows a status bar with various keyboard shortcuts: ^G Get Help, ^O Write Out, ^W Where Is, ^K Cut Text, ^J Justify, ^C Cur Pos, ^X Exit, ^R Read File, ^\ Replace, ^U Uncut Text, ^T To Spell, ^\_ Go To Line.



# HANDS-ON INSTALLING HADOOP

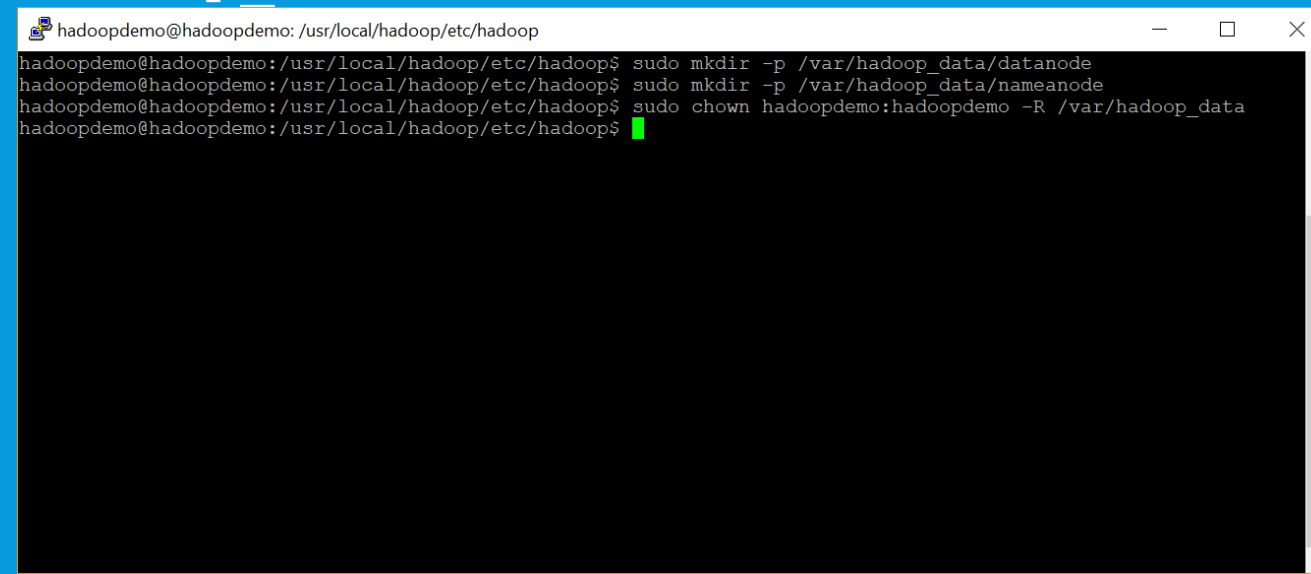
## 6. Configure Hadoop

- 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
```

A terminal window with a black background and white text. The window title is 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. It shows three commands being executed: 'sudo mkdir -p /var/hadoop\_data/datanode', 'sudo mkdir -p /var/hadoop\_data/nameanode', and 'sudo chown hadoopdemo:hadoopdemo -R /var/hadoop\_data'. Each command is followed by a green cursor. The window has standard Linux window controls (minimize, maximize, close) in the top right corner.

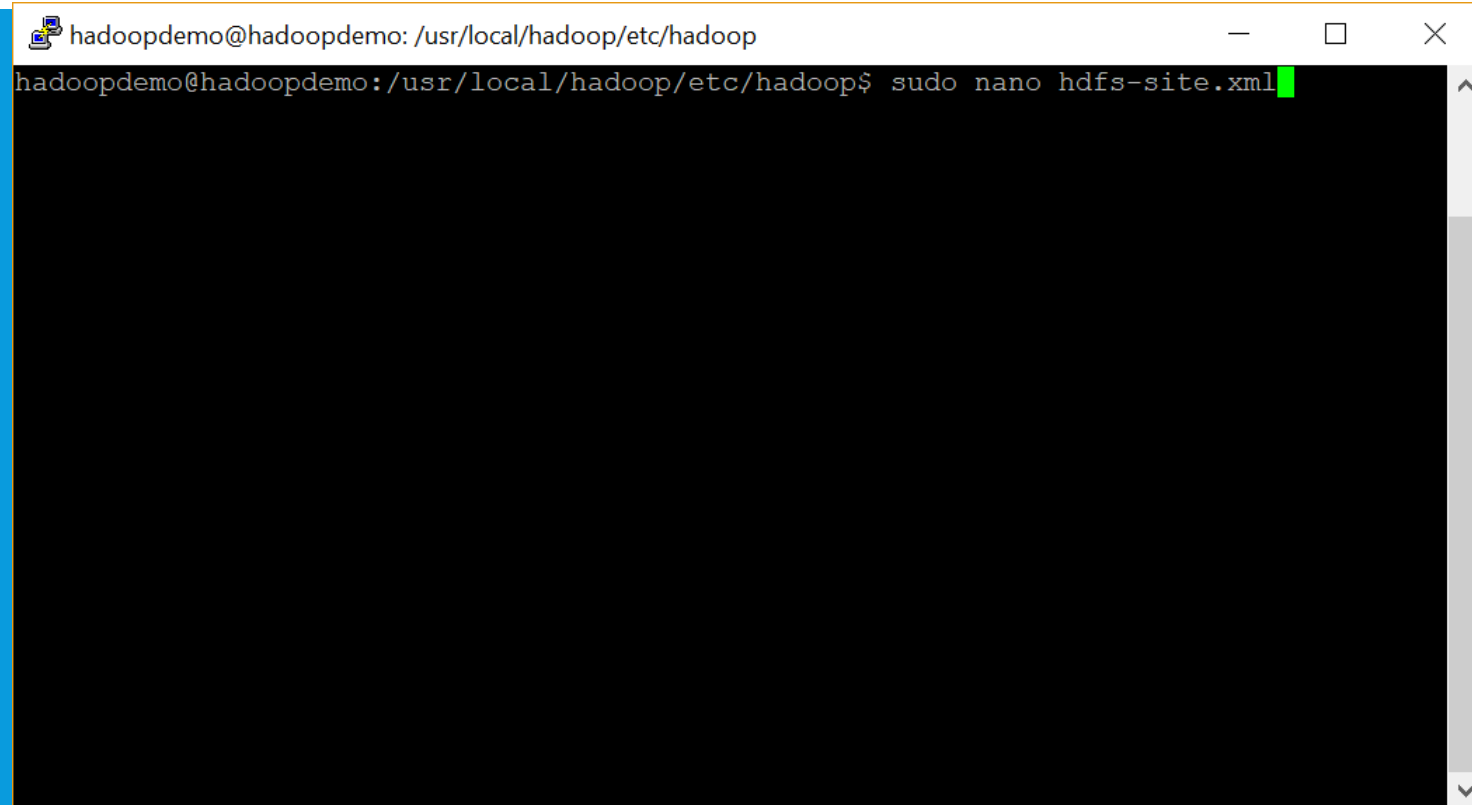
```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /var/hadoop_data/datanode
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo mkdir -p /var/hadoop_data/nameanode
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo chown hadoopdemo:hadoopdemo -R /var/hadoop_data
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$
```

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Edit hdfs-site.xml

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

A terminal window with a black background and white text. The window title bar shows 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The command 'sudo nano hdfs-site.xml' has been entered and executed, as indicated by a green cursor at the end of the line.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop  
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ 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>
```

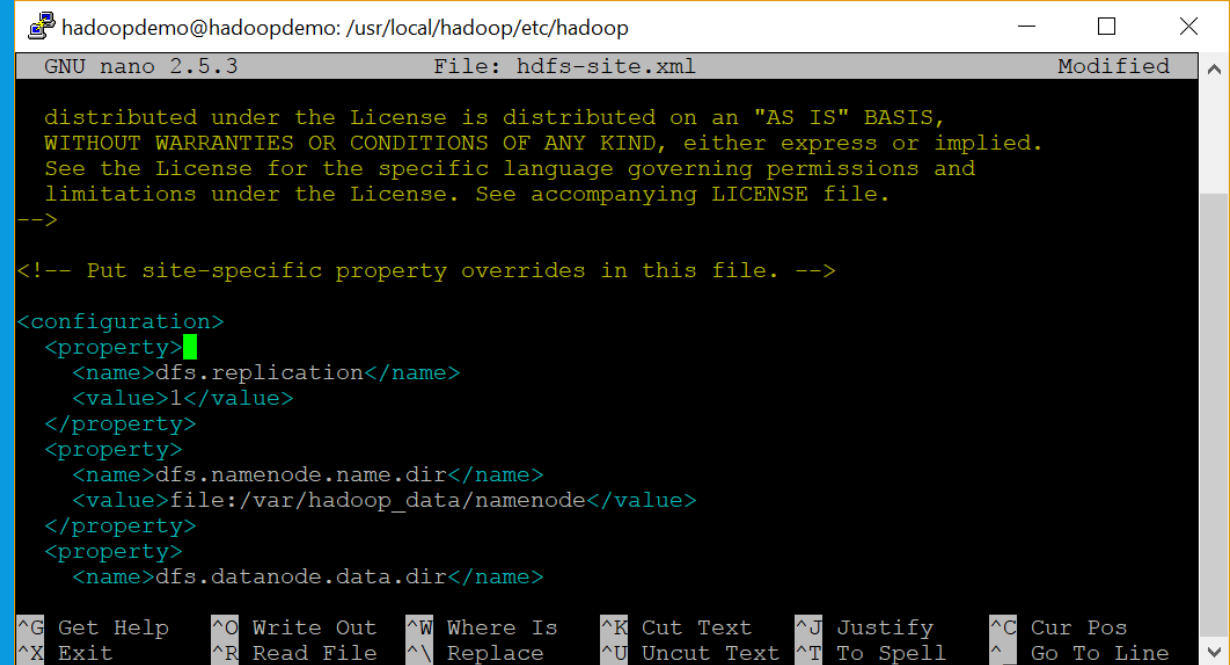
```
<property>
```

```
<name>dfs.datanode.data.dir</name>
```

```
<value>file:/var/hadoop_data/datanode</value>
```

```
</property>
```

```
</configuration>
```



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
GNU nano 2.5.3 File: hdfs-site.xml Modified

distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

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

<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>

```

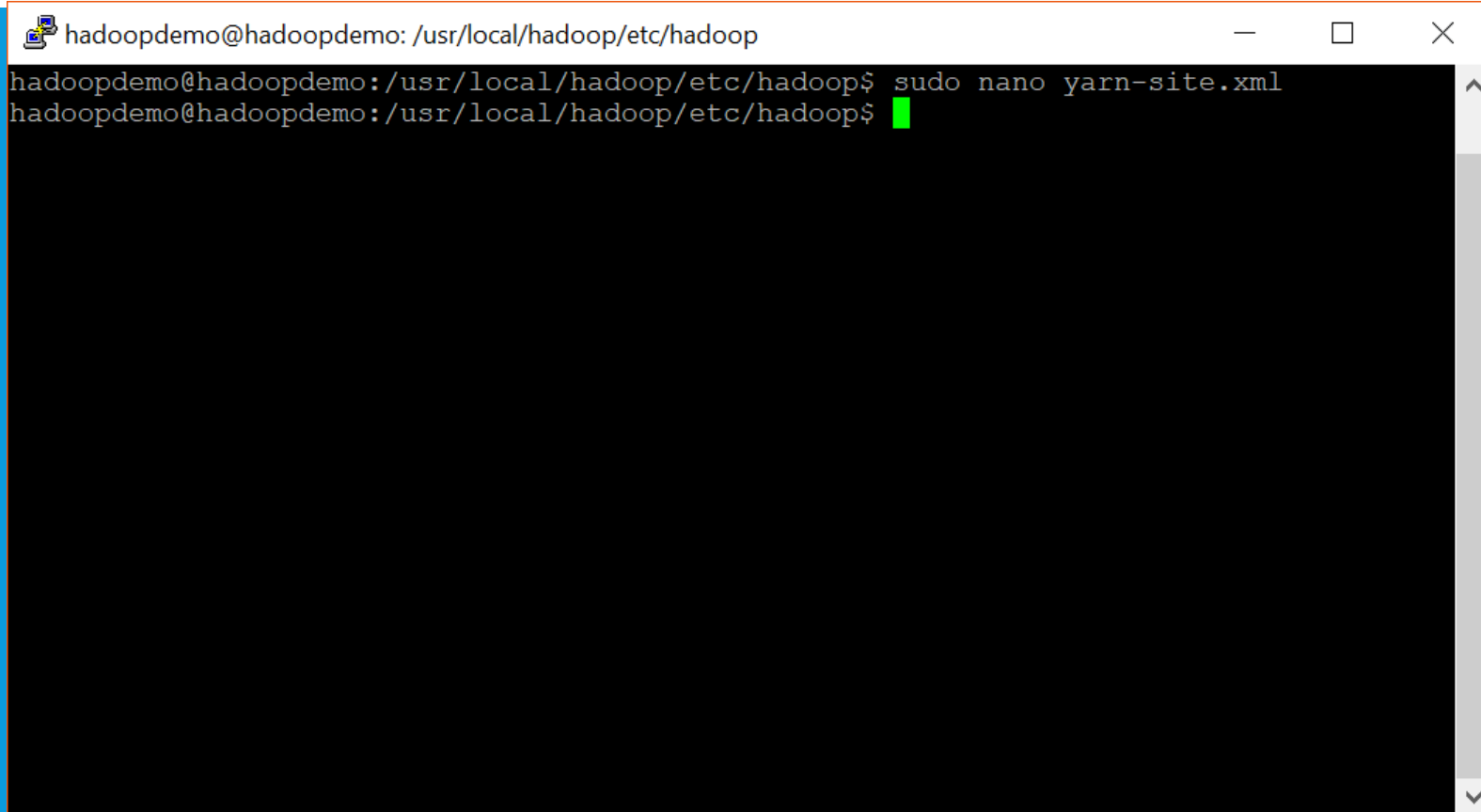
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell ^\_ Go To Line

# HANDS-ON INSTALLING HADOOP

## 6. Configure Hadoop

- Edit yarn-site.xml

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

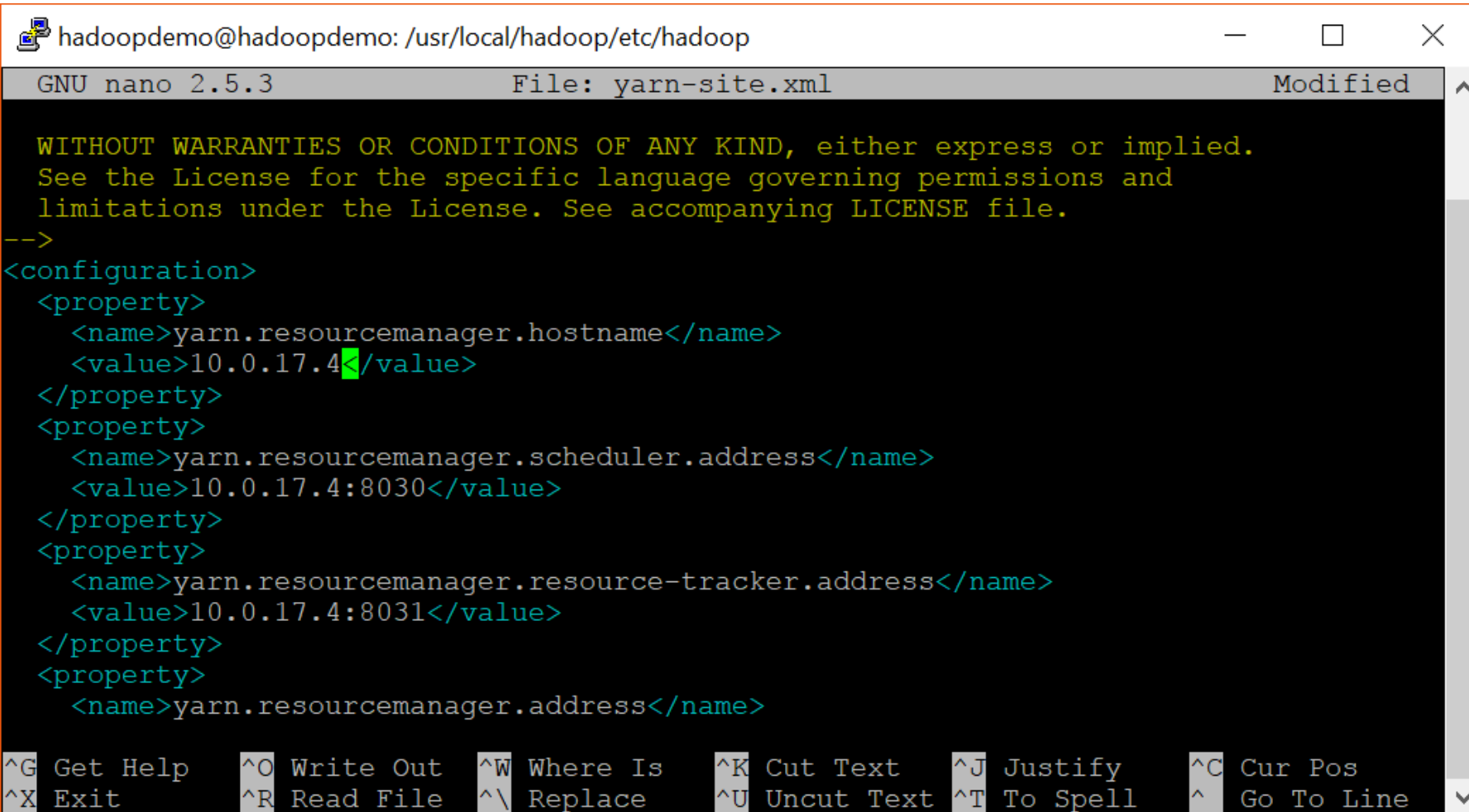
A terminal window with a black background and white text. The window title is 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The prompt is 'hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop\$'. The command 'sudo nano yarn-site.xml' has been entered and executed, resulting in a new prompt 'hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop\$' with a green cursor. The terminal window has standard Linux window controls (minimize, maximize, close) in the top right corner.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo nano yarn-site.xml
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$
```

# 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



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
GNU nano 2.5.3 File: yarn-site.xml Modified
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<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></value>
  </property>

```

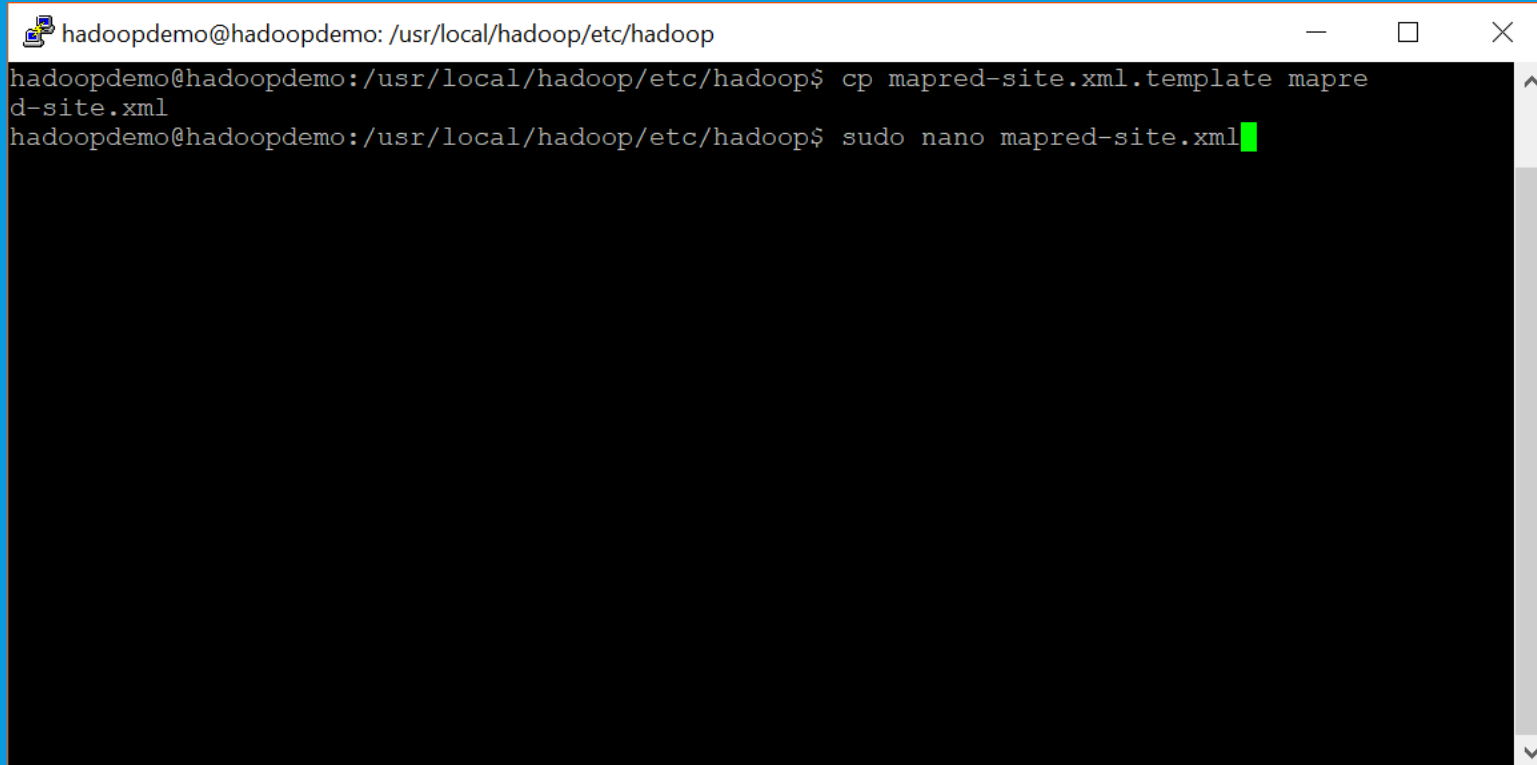
<b>^G</b> Get Help	<b>^O</b> Write Out	<b>^W</b> Where Is	<b>^K</b> Cut Text	<b>^J</b> Justify	<b>^C</b> Cur Pos
<b>^X</b> Exit	<b>^R</b> Read File	<b>^_</b> Replace	<b>^U</b> Uncut Text	<b>^T</b> To Spell	<b>^_</b> Go To Line

# HANDS-ON INSTALLING HADOOP

## 6.Configure Hadoop

- Edit mapred-site.xml

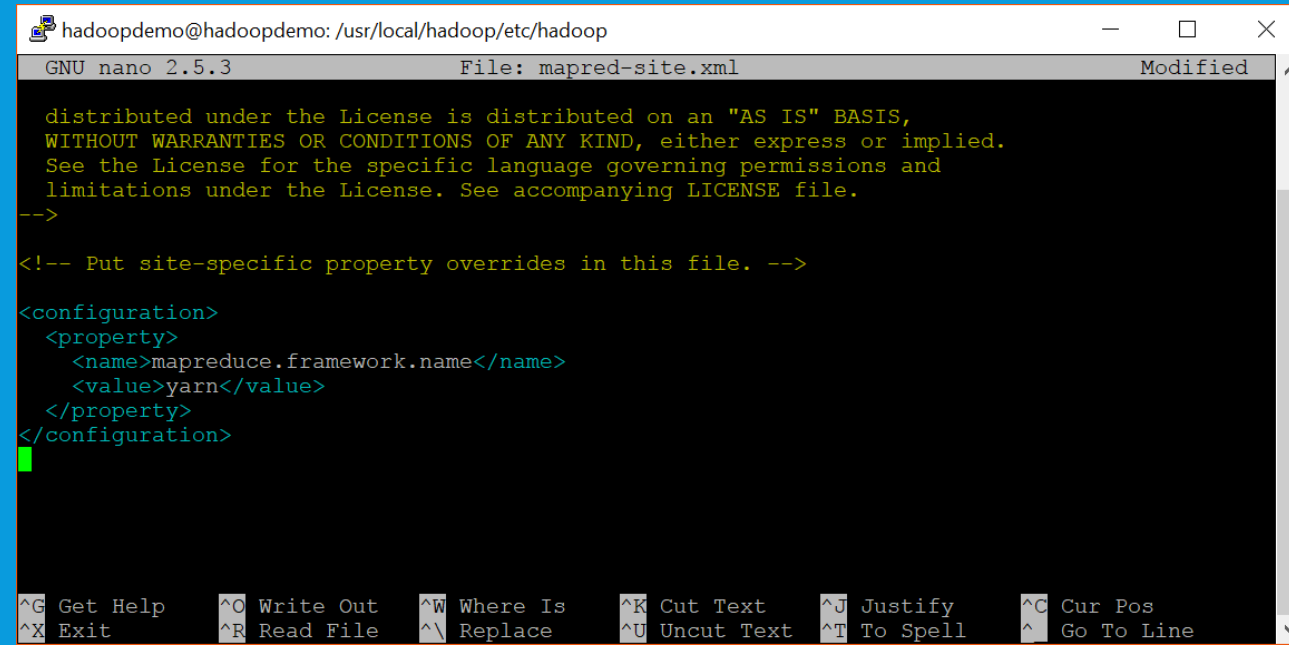
*\$ sudo nano mapred-site.xml*

A terminal window with a title bar showing 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The terminal contains two lines of commands: 'cp mapred-site.xml.template mapred-site.xml' and 'sudo nano mapred-site.xml'. The second command is followed by a green cursor. The terminal has a dark background and a light-colored title bar with standard window controls.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ cp mapred-site.xml.template mapred-site.xml
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ sudo nano mapred-site.xml
```

# HANDS-ON INSTALLING HADOOP

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



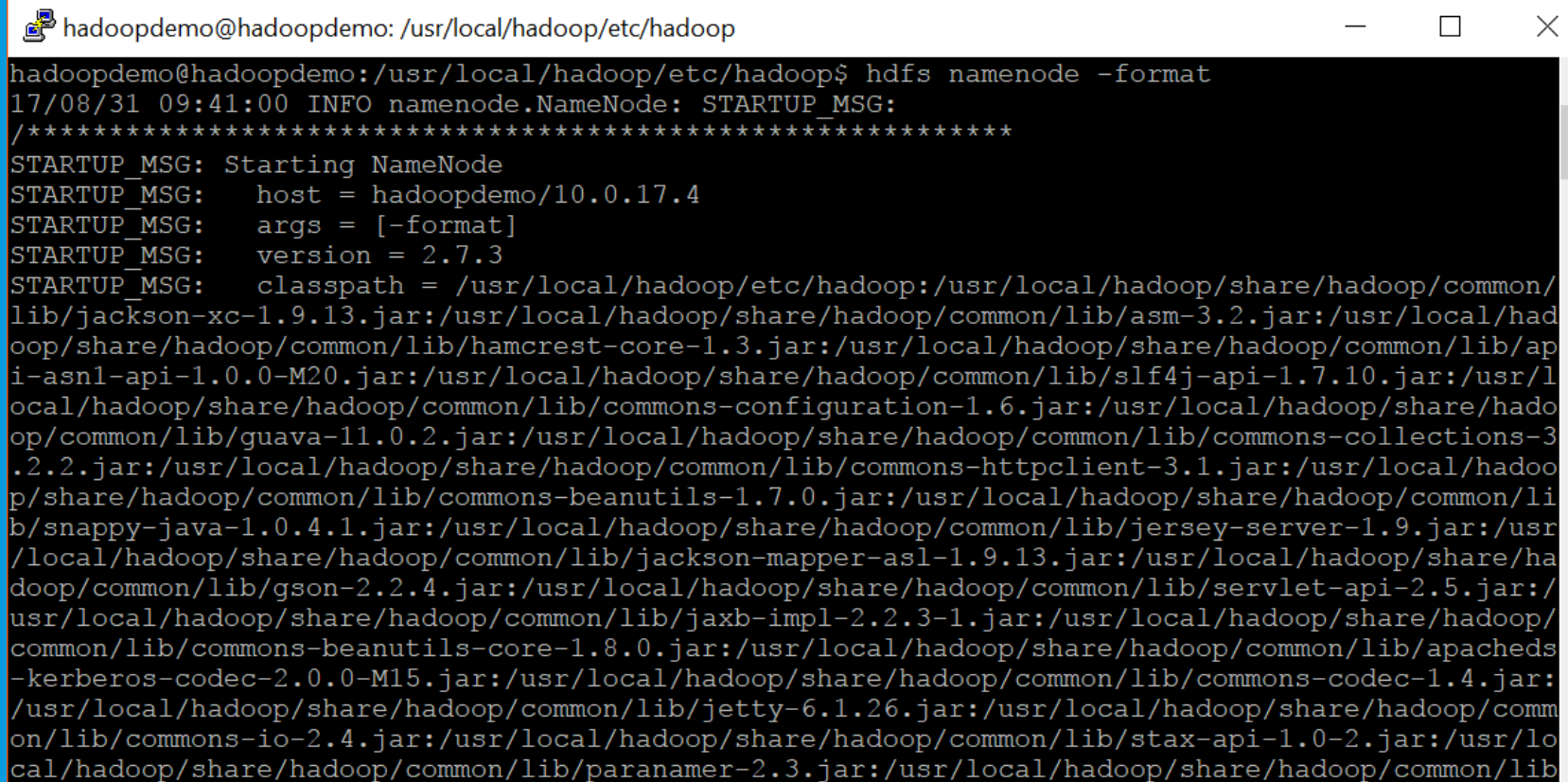
```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
GNU nano 2.5.3 File: mapred-site.xml Modified
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Uncut Text ^T To Spell ^_ Go To Line
```



# HANDS-ON INSTALLING HADOOP

## 7.Formatting Namenode

```
$ hdfs namenode -format
```

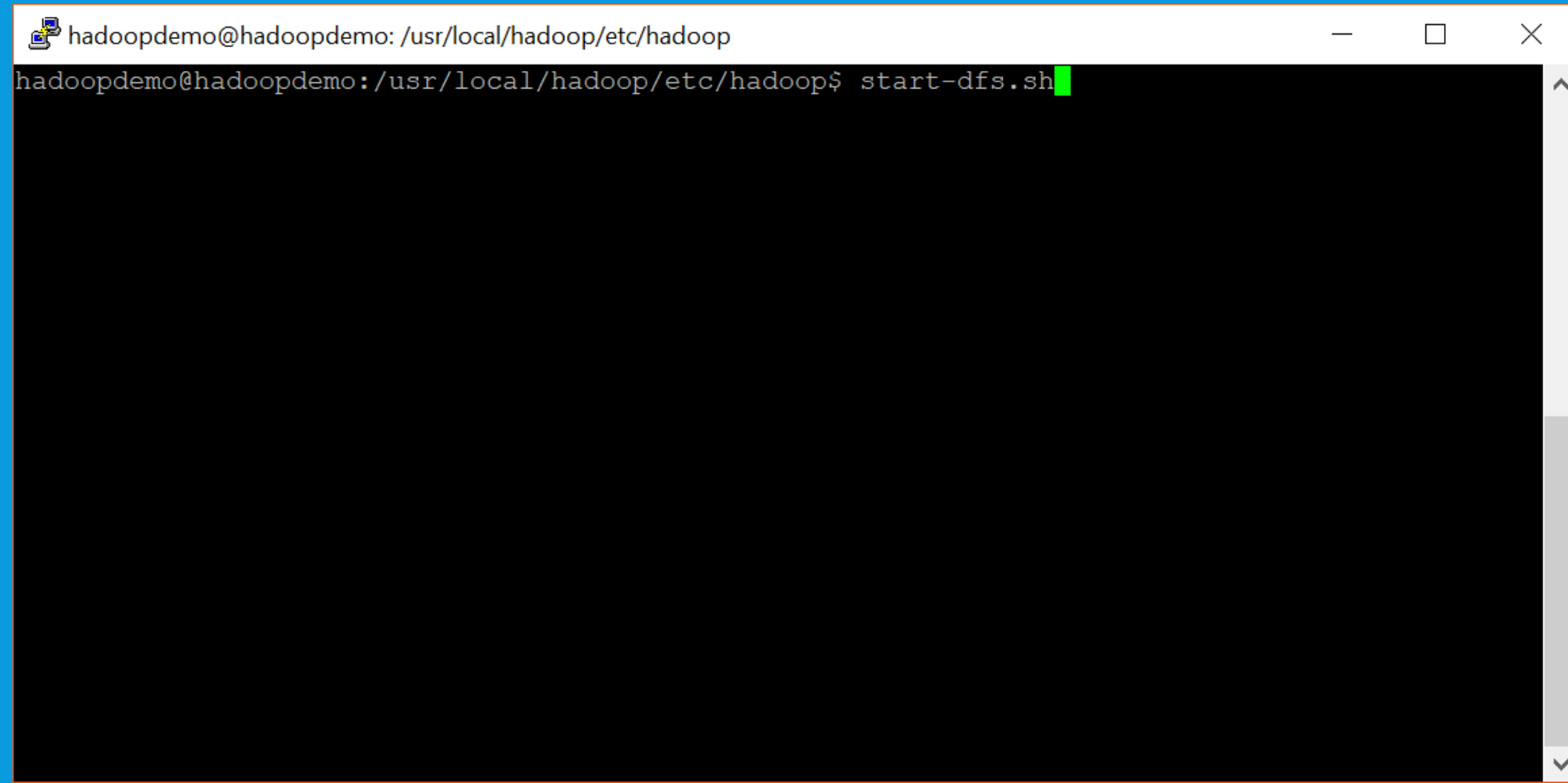


```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ hdfs namenode -format
17/08/31 09:41:00 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:  host = hadoopdemo/10.0.17.4
STARTUP_MSG:  args = [-format]
STARTUP_MSG:  version = 2.7.3
STARTUP_MSG:  classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/
lib/jackson-xc-1.9.13.jar:/usr/local/hadoop/share/hadoop/common/lib/asm-3.2.jar:/usr/local/had
oop/share/hadoop/common/lib/hamcrest-core-1.3.jar:/usr/local/hadoop/share/hadoop/common/lib/ap
i-asn1-api-1.0.0-M20.jar:/usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.10.jar:/usr/l
ocal/hadoop/share/hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/hadoop/share/hado
op/common/lib/guava-11.0.2.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-collections-3
.2.2.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-httpclient-3.1.jar:/usr/local/hadoo
p/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/usr/local/hadoop/share/hadoop/common/li
b/snappy-java-1.0.4.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-server-1.9.jar:/usr
/local/hadoop/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/hadoop/share/ha
dooop/common/lib/gson-2.2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/servlet-api-2.5.jar:/
usr/local/hadoop/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/usr/local/hadoop/share/hadoop/
common/lib/commons-beanutils-core-1.8.0.jar:/usr/local/hadoop/share/hadoop/common/lib/apacheds
-kerberos-codec-2.0.0-M15.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-codec-1.4.jar:
/usr/local/hadoop/share/hadoop/common/lib/jetty-6.1.26.jar:/usr/local/hadoop/share/hadoop/comm
on/lib/commons-io-2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/stax-api-1.0-2.jar:/usr/lo
cal/hadoop/share/hadoop/common/lib/paranamer-2.3.jar:/usr/local/hadoop/share/hadoop/common/lib
```

# HANDS-ON INSTALLING HADOOP

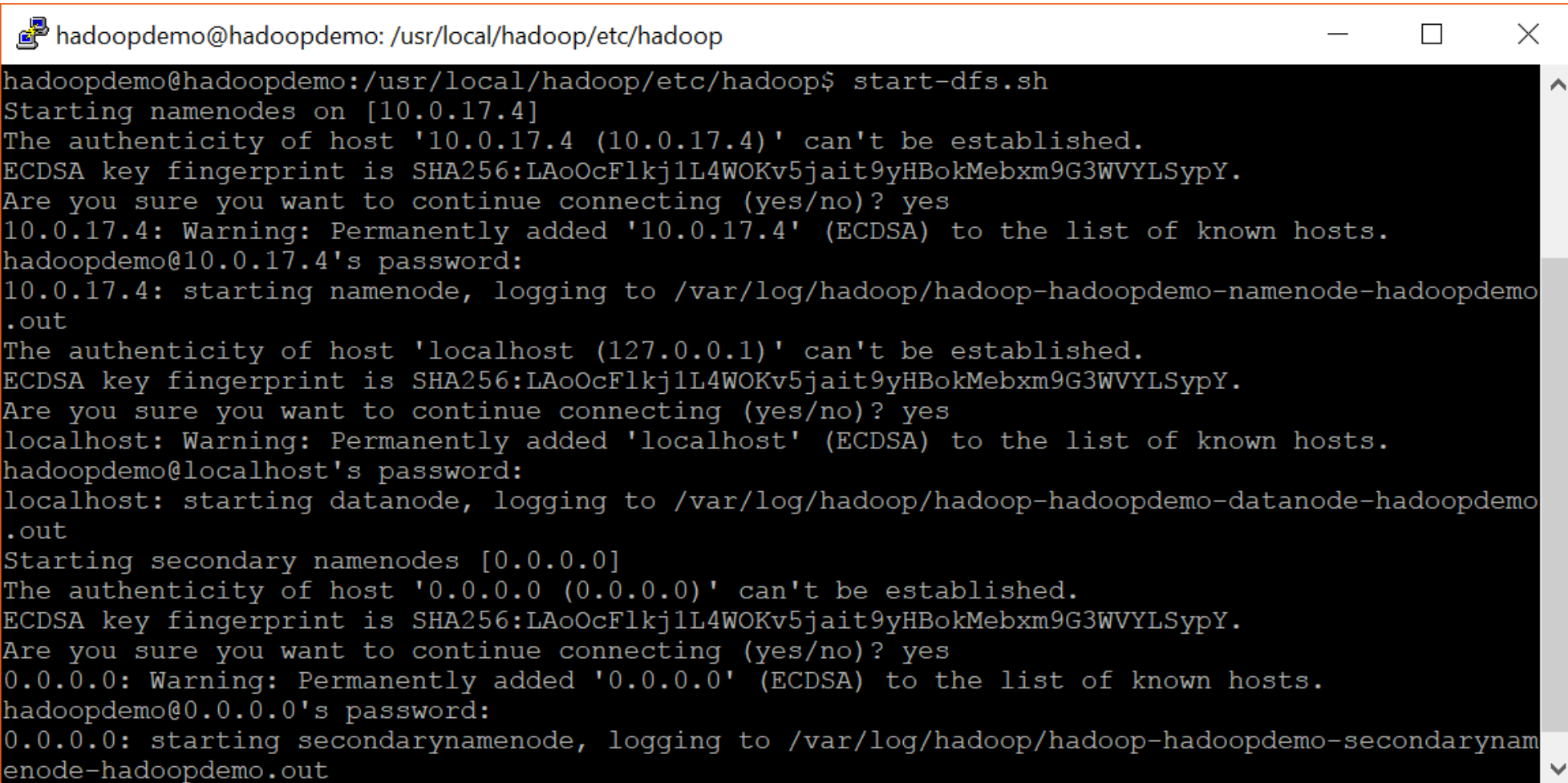
## ▪ 8.Starting Namenode and Datanode

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

A terminal window with a white title bar and standard window controls (minimize, maximize, close). The title bar text is 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The terminal content shows the command 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop\$ start-dfs.sh' followed by a green cursor. The rest of the terminal area is black and empty.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop$ start-dfs.sh
```

# HANDS-ON INSTALLING HADOOP

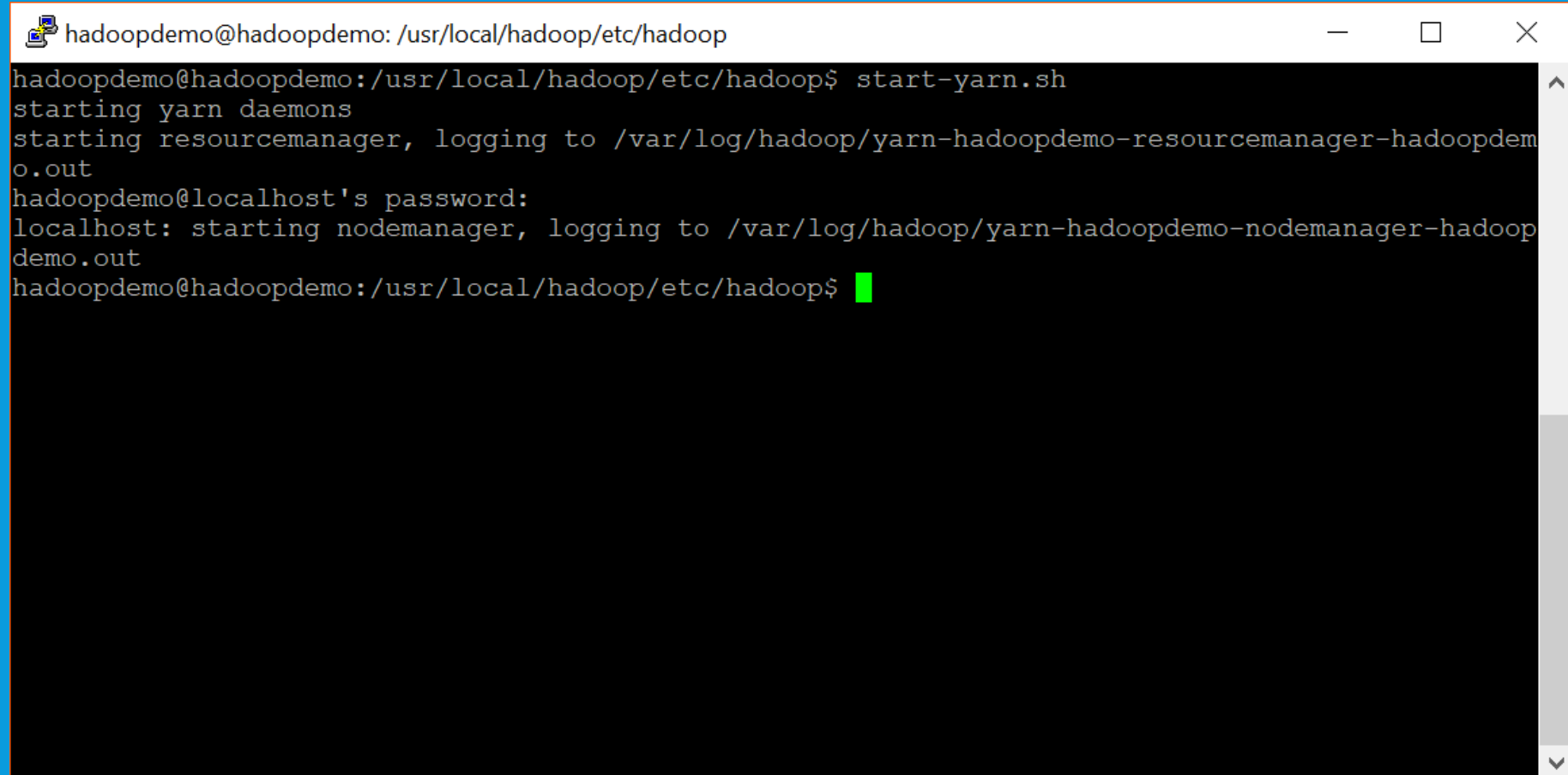
A terminal window with a title bar showing the user 'hadoopdemo' at 'hadoopdemo' in the directory '/usr/local/hadoop/etc/hadoop'. The terminal output shows the execution of 'start-dfs.sh', which starts namenodes on '10.0.17.4'. It prompts for a password and shows a warning about adding the host to the list of known hosts. Then it starts a datanode on 'localhost' (127.0.0.1), also prompting for a password and showing a warning. Finally, it starts secondary namenodes on '0.0.0.0', again prompting for a password and showing a warning. The terminal text is as follows:

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ start-dfs.sh
Starting namenodes on [10.0.17.4]
The authenticity of host '10.0.17.4 (10.0.17.4)' can't be established.
ECDSA key fingerprint is SHA256:LAoOcFlkjlL4WOKv5jait9yHBokMebxm9G3WVYLSypY.
Are you sure you want to continue connecting (yes/no)? yes
10.0.17.4: Warning: Permanently added '10.0.17.4' (ECDSA) to the list of known hosts.
hadoopdemo@10.0.17.4's password:
10.0.17.4: starting namenode, logging to /var/log/hadoop/hadoop-hadoopdemo-namenode-hadoopdemo.out
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:LAoOcFlkjlL4WOKv5jait9yHBokMebxm9G3WVYLSypY.
Are you sure you want to continue connecting (yes/no)? yes
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
hadoopdemo@localhost's password:
localhost: starting datanode, logging to /var/log/hadoop/hadoop-hadoopdemo-datanode-hadoopdemo.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:LAoOcFlkjlL4WOKv5jait9yHBokMebxm9G3WVYLSypY.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.
hadoopdemo@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /var/log/hadoop/hadoop-hadoopdemo-secondarynamenode-hadoopdemo.out
```

# HANDS-ON INSTALLING HADOOP

## 8.Starting Yarn

*\$ start-yarn.sh*

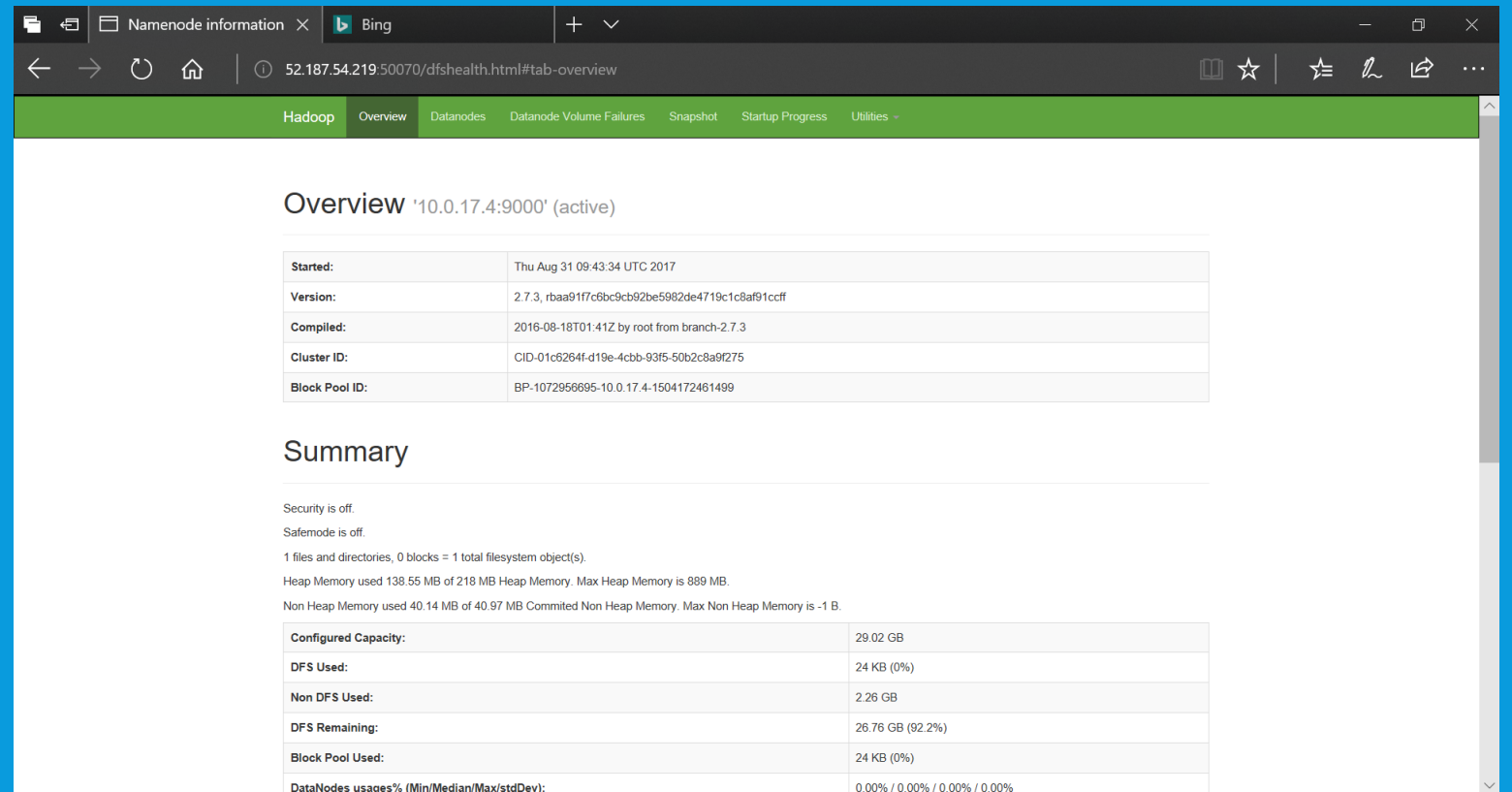
A terminal window with a title bar showing 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop'. The terminal output shows the command 'start-yarn.sh' being executed, followed by 'starting yarn daemons', 'starting resourcemanager, logging to /var/log/hadoop/yarn-hadoopdemo-resourcemanager-hadoopdemo.out', a password prompt 'hadoopdemo@localhost's password:', and 'localhost: starting nodemanager, logging to /var/log/hadoop/yarn-hadoopdemo-nodemanager-hadoopdemo.out'. The prompt returns to 'hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop\$' with a green cursor.

```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /var/log/hadoop/yarn-hadoopdemo-resourcemanager-hadoopdemo.out
hadoopdemo@localhost's password:
localhost: starting nodemanager, logging to /var/log/hadoop/yarn-hadoopdemo-nodemanager-hadoopdemo.out
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop$
```

# HANDS-ON INSTALLING HADOOP

## 9. Accessing Hadoop Web Console

- My public ip is 52.187.54.219



Namenode information X Bing

52.187.54.219:50070/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Overview '10.0.17.4:9000' (active)

Started:	Thu Aug 31 09:43:34 UTC 2017
Version:	2.7.3, rbaa91f7c6bc9cb92be5982de4719c1c8af91ccff
Compiled:	2016-08-18T01:41Z by root from branch-2.7.3
Cluster ID:	CID-01c6264f-d19e-4cbb-93f5-50b2c8a9f275
Block Pool ID:	BP-1072956695-10.0.17.4-1504172461499

### Summary

Security is off.  
Safemode is off.  
1 files and directories, 0 blocks = 1 total filesystem object(s).  
Heap Memory used 138.55 MB of 218 MB Heap Memory. Max Heap Memory is 889 MB.  
Non Heap Memory used 40.14 MB of 40.97 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

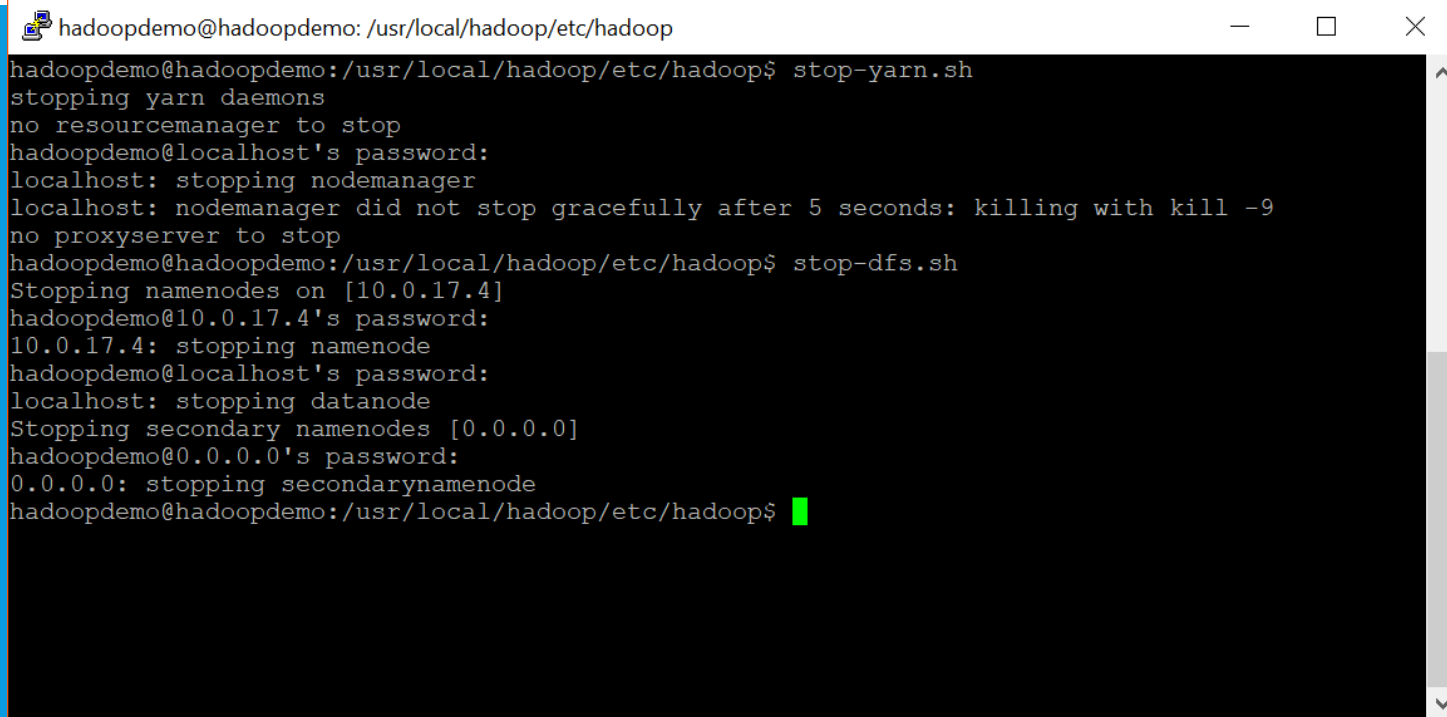
Configured Capacity:	29.02 GB
DFS Used:	24 KB (0%)
Non DFS Used:	2.26 GB
DFS Remaining:	26.76 GB (92.2%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%

# HANDS-ON INSTALLING HADOOP

## 10. Stop Hadoop Yarn and HDFS

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

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



```
hadoopdemo@hadoopdemo: /usr/local/hadoop/etc/hadoop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ stop-yarn.sh
stopping yarn daemons
no resourcemanager to stop
hadoopdemo@localhost's password:
localhost: stopping nodemanager
localhost: nodemanager did not stop gracefully after 5 seconds: killing with kill -9
no proxyserver to stop
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$ stop-dfs.sh
Stopping namenodes on [10.0.17.4]
hadoopdemo@10.0.17.4's password:
10.0.17.4: stopping namenode
hadoopdemo@localhost's password:
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
hadoopdemo@0.0.0.0's password:
0.0.0.0: stopping secondarynamenode
hadoopdemo@hadoopdemo:/usr/local/hadoop/etc/hadoop$
```

# HANDS-ON: TRAVERSING, RETRIEVING DATA FROM HDFS

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

```
$ hdfs dfs -mkdir /inputs
```

```
$ hdfs dfs -mkdir /outputs
```

Get data file from internet

```
$ wget
```

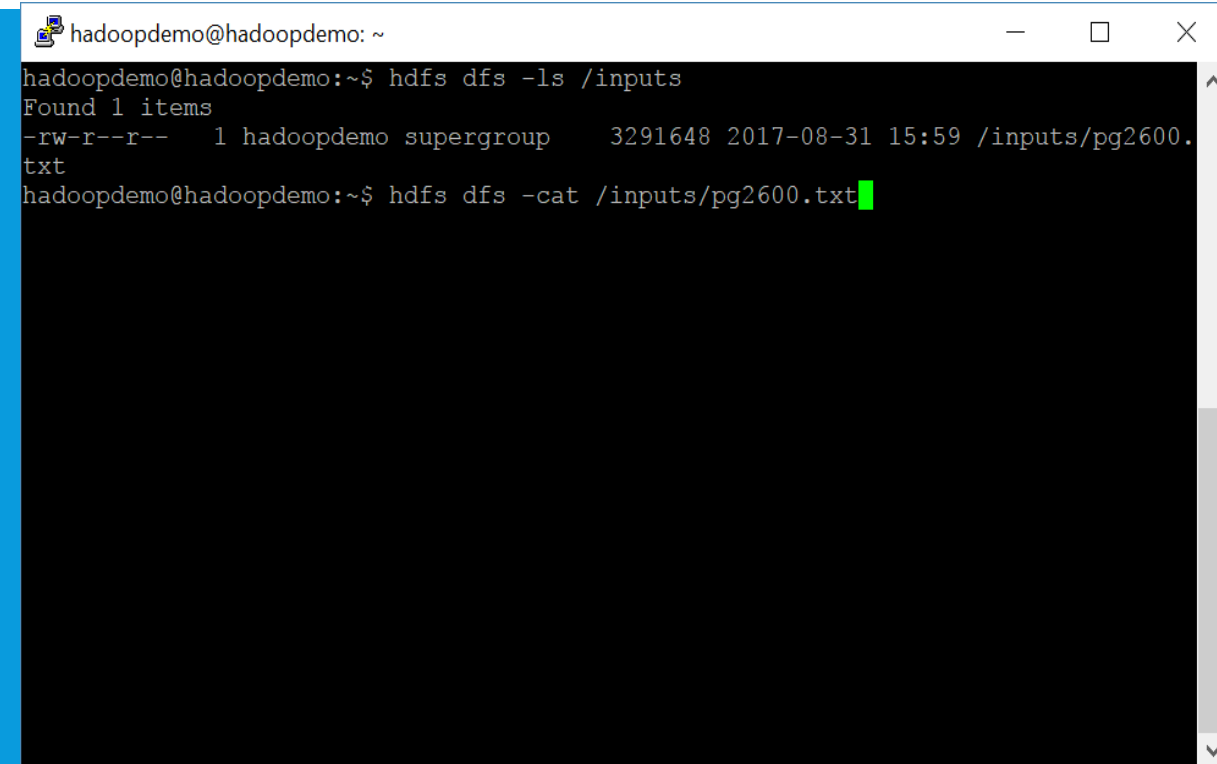
```
https://s3.amazonaws.com/imcbucket/  
input/pg2600.txt
```

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

```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ hdfs dfs -mkdir /inputs  
hadoopdemo@hadoopdemo:~$ hdfs dfs -mkdir /outputs  
hadoopdemo@hadoopdemo:~$ wget https://s3.amazonaws.com/imcbucket/input/pg2600.tx  
t  
--2017-08-31 15:49:28-- https://s3.amazonaws.com/imcbucket/input/pg2600.txt  
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.216.1.195  
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.216.1.195|:443... connected  
.  
HTTP request sent, awaiting response... 200 OK  
Length: 3291648 (3.1M) [text/plain]  
Saving to: 'pg2600.txt.1'  
  
pg2600.txt.1      100%[=====>]    3.14M  1.16MB/s   in 2.7s  
2017-08-31 15:49:32 (1.16 MB/s) - 'pg2600.txt.1' saved [3291648/3291648]  
hadoopdemo@hadoopdemo:~$
```

# HANDS-ON: TRAVERSING, RETRIEVING DATA FROM HDFS

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

A terminal window titled 'hadoopdemo@hadoopdemo: ~' with standard window controls. It shows the execution of HDFS commands. The first command 'hdfs dfs -ls /inputs' returns 'Found 1 items' followed by a file listing: '-rw-r--r-- 1 hadoopdemo supergroup 3291648 2017-08-31 15:59 /inputs/pg2600.txt'. The second command 'hdfs dfs -cat /inputs/pg2600.txt' is entered, and a green cursor is visible at the end of the line.

```
hadoopdemo@hadoopdemo: ~  
hadoopdemo@hadoopdemo:~$ hdfs dfs -ls /inputs  
Found 1 items  
-rw-r--r-- 1 hadoopdemo supergroup 3291648 2017-08-31 15:59 /inputs/pg2600.  
txt  
hadoopdemo@hadoopdemo:~$ hdfs dfs -cat /inputs/pg2600.txt
```