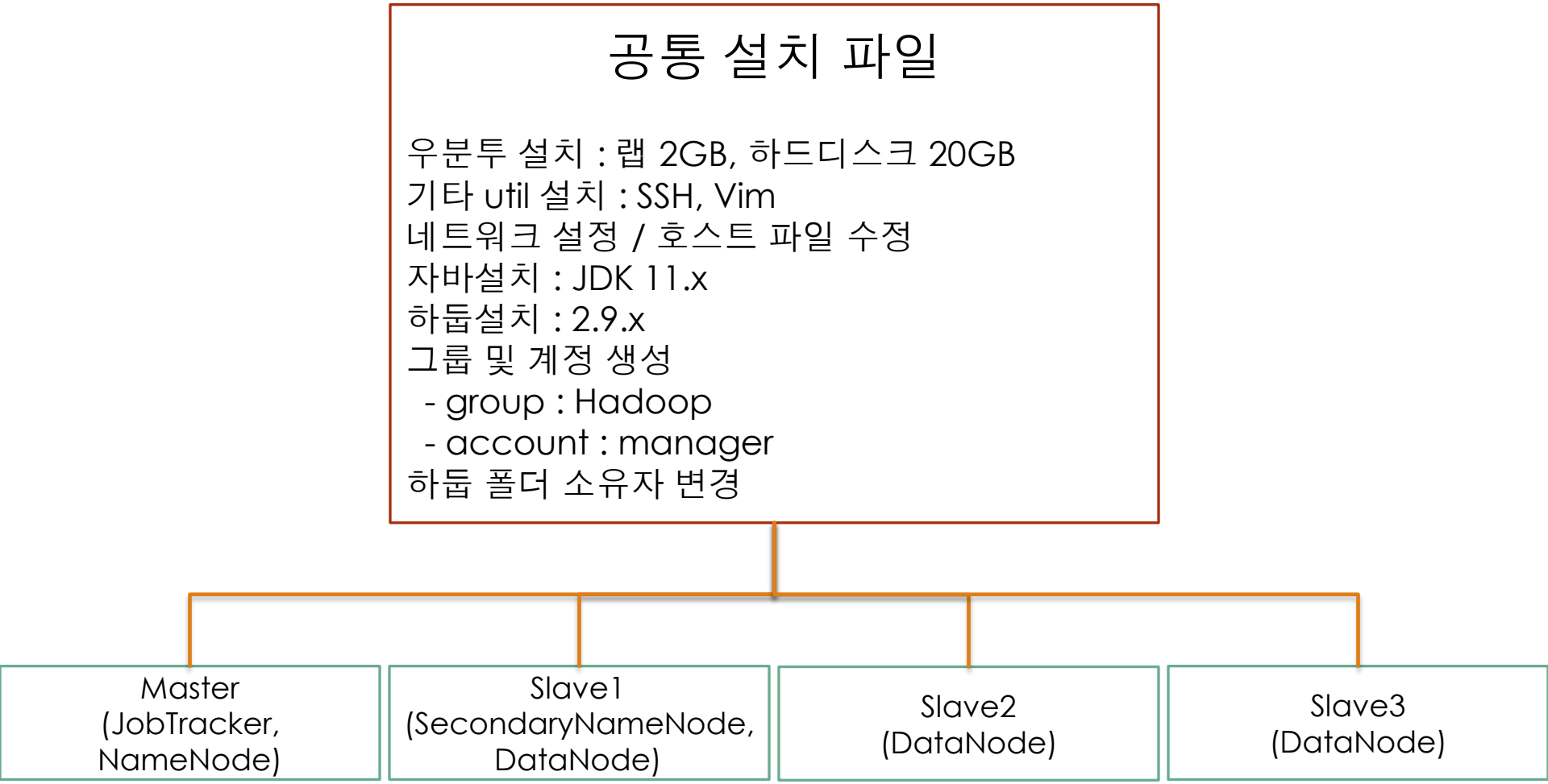
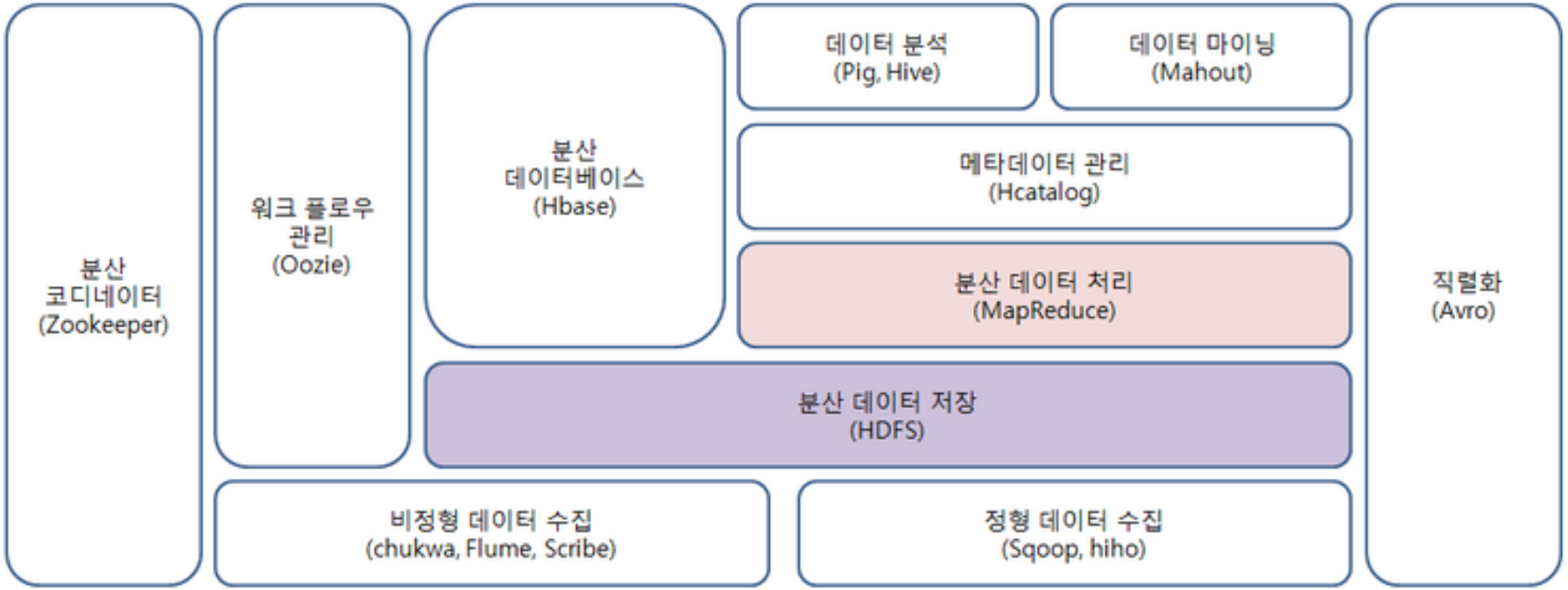
완전 분산 모드

환경구성

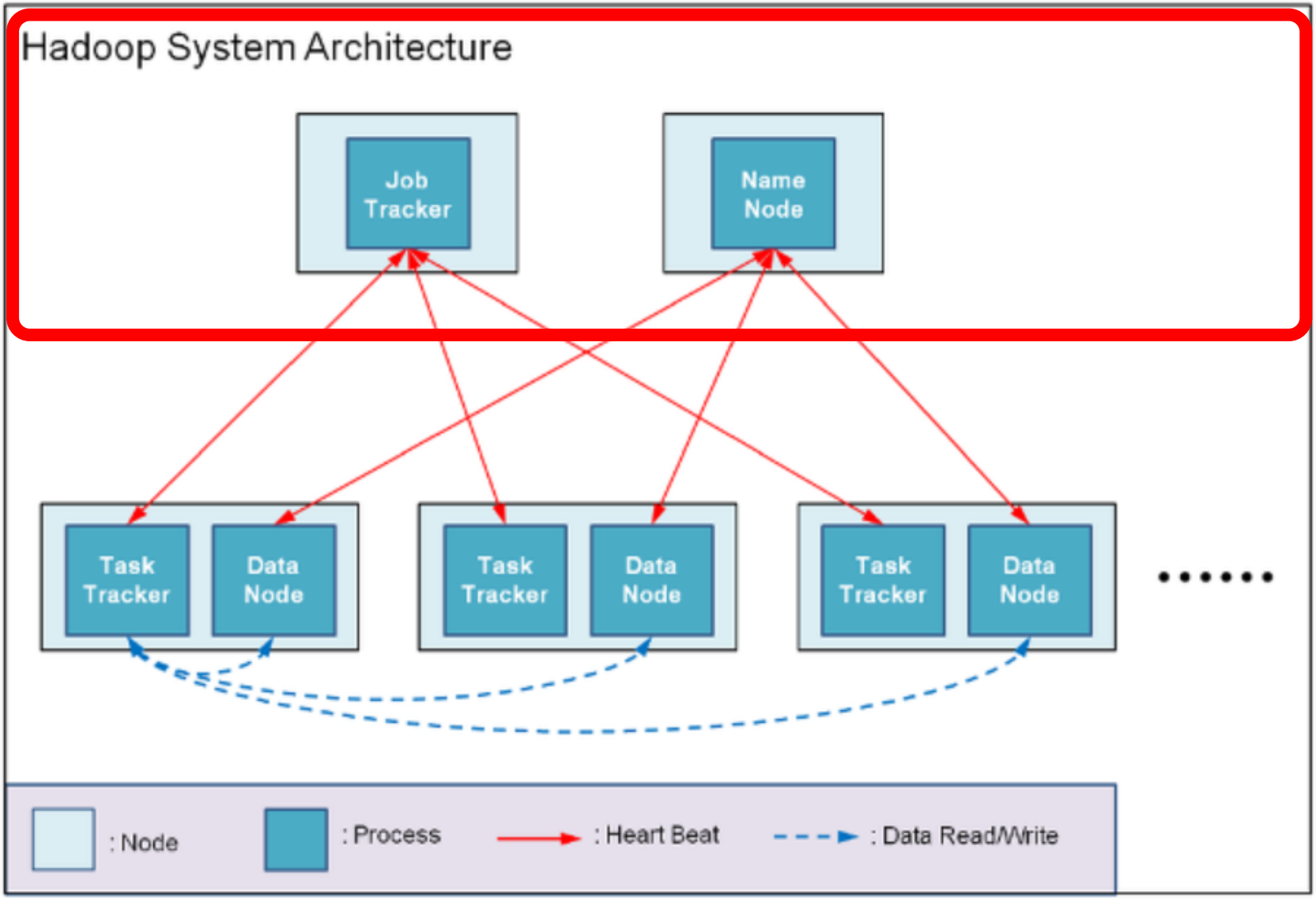
구성도


공통 설치 파일
우분투 설치 : 랩 2GB, 하드디스크 20GB
기타 util 설치 : SSH, Vim
네트워크 설정 / 호스트 파일 수정
자바설치 : JDK 11.x
하둡설치 : 2.9.x
그룹 및 계정 생성
- group : Hadoop
- account : manager
하둡 폴더 소유자 변경
Master (JobTracker, NameNode)
Slave1 (SecondaryNameNode, DataNode)
Slave2 (DataNode)
Slave3 (DataNode)

Hadoop 구성


분산
코디네이터
(Zookeeper)
워크 플로우
관리
(Oozie)
분산
데이터베이스
(Hbase)
데이터 분석
(Pig, Hive)
데이터 마이닝
(Mahout)
메타데이터 관리
(Hcatalog)
분산 데이터 처리
(MapReduce)
분산 데이터 저장
(HDFS)
직렬화
(Avro)
비정형 데이터 수집
(chukwa, Flume, Scribe)
정형 데이터 수집
(Sqoop, hiho)

우분투 설치


Hadoop System Architecture
Job Tracker
Name Node
Task Tracker
Data Node
Task Tracker
Data Node
Task Tracker
Data Node
: Node
: Process
: Heart Beat
: Data Read/Write

Master 설정

하둡 초기화

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

```
$ rm -rf /usr/local/hadoop/hdfs/datanode/
$ rm -rf /usr/local/hadoop/hdfs/namenode/*
```

Hosts 파일 변경

```
manager@jin-VirtualBox:~$ sudo vim /etc/hosts
[sudo] manager의  암 호 :
```

```
127.0.0.1       localhost
127.0.1.1       jin-VirtualBox

192.168.56.101  master
192.168.56.102  slave1
192.168.56.103  slave2
192.168.56.104  slave3

# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
```

core-site 수정

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

```
<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/usr/local/hadoop/tmp</value>
        </property>

        <property>
                <name>fs.default.name</name>
                <value>hdfs://master:62350</value>
        </property>
</configuration>
```

mapred-site 수정 / Hadoop 설정파일 저장

```
$ sudo vim /usr/local/hadoop/etc/hadoop/mapred-site.xml
```

```
<configuration>
        <property>
                <name>mapred.job.tracker</name>
                <value>master:62351</value>
        </property>
</configuration>
```

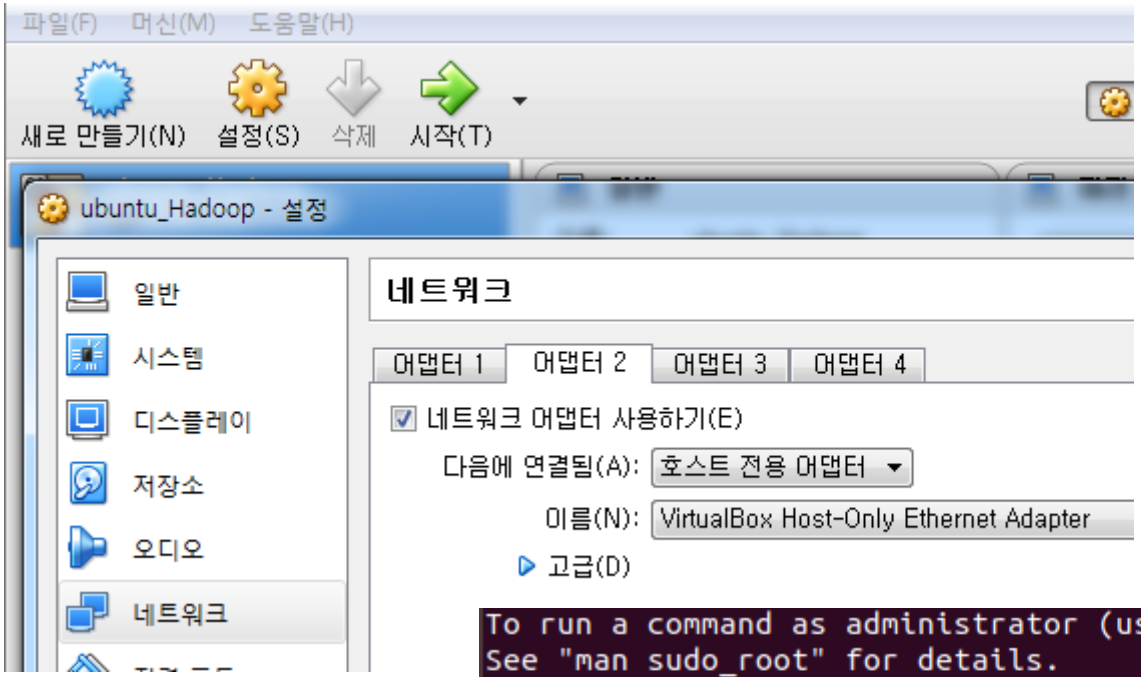
```
$ tar -cvf hadoop.tar /usr/local/hadoop
```

프롬프트 변경

```
jin@jin-VirtualBox:~$ sudo vim /etc/hostname
[sudo] jin의  암호 :
jin@jin-VirtualBox:~$ sudo /bin/hostname -F /etc/hostname
jin@jin-VirtualBox:~$ sudo reboot
```

Slave 만들기

네트워크 설정



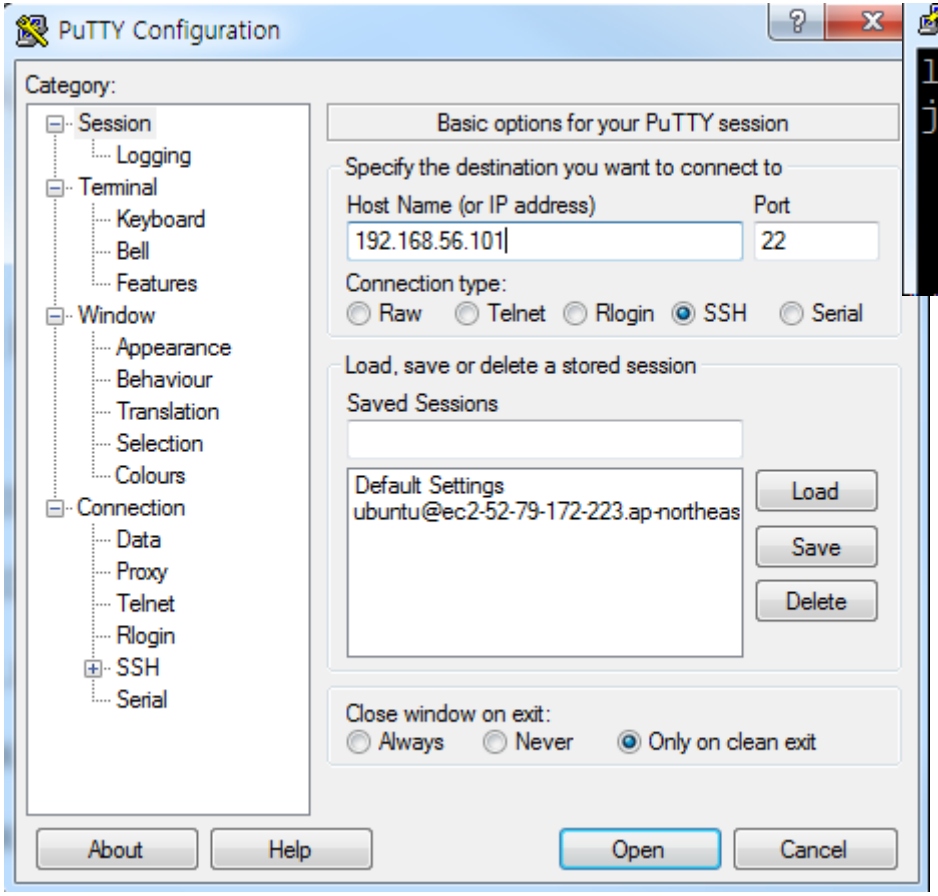

```
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

jin@jin-VirtualBox:~$ ifconfig

Command 'ifconfig' not found, but can be installed with:

sudo apt install net-tools

jin@jin-VirtualBox:~$ sudo apt install net-tools
```

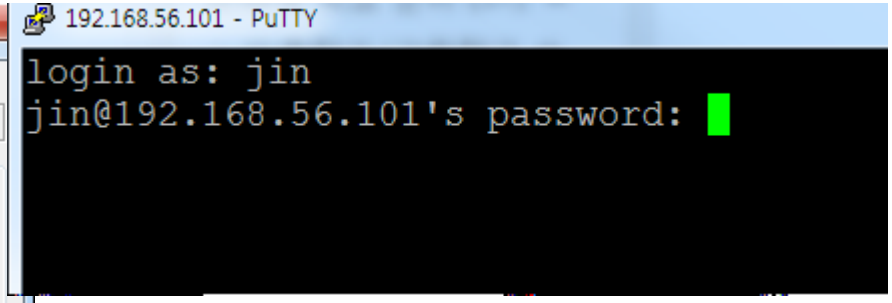
유틸 설치

```
jin@jin-VirtualBox:~$ sudo apt-get install ssh
패키지 목록을 읽는 중입니다... 완료
의존성 트리를 만드는 중입니다
상태 정보를 읽는 중입니다... 완료
The following additional packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
제안하는 패키지:
  ssh-askpass rssh molly-guard monkeysphere
다음 새 패키지를 설치할 것입니다:
```

```
jin@jin-VirtualBox:~$ sudo apt-get install vim -y
```

Putty 접속


PuTTY Configuration
Category:
Session
Logging
Terminal
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
Telnet
Rlogin
SSH
Serial
Basic options for your PuTTY session
Specify the destination you want to connect to
Host Name (or IP address)
Port
192.168.56.101
22
Connection type:
Raw
Telnet
Rlogin
SSH
Serial
Load, save or delete a stored session
Saved Sessions
Default Settings
ubuntu@ec2-52-79-172-223.ap-northeas
Load
Save
Delete
Close window on exit:
Always
Never
Only on clean exit
About
Help
Open
Cancel


192.168.56.101 - PuTTY
login as: jin
jin@192.168.56.101's password:

JDK 설치

```
jin@jin-VirtualBox:~$ sudo apt-get update
```

```
jin@jin-VirtualBox:~$ java -version
```

```
jin@jin-VirtualBox:~$ sudo apt-get install default-jdk -y
```

```
jin@jin-VirtualBox:~$ which java
/usr/bin/java
jin@jin-VirtualBox:~$ readlink -f /usr/bin/java
/usr/lib/jvm/java-11-openjdk-amd64/bin/java
```

```
jin@jin-VirtualBox:~$ sudo vim /etc/profile
```

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64

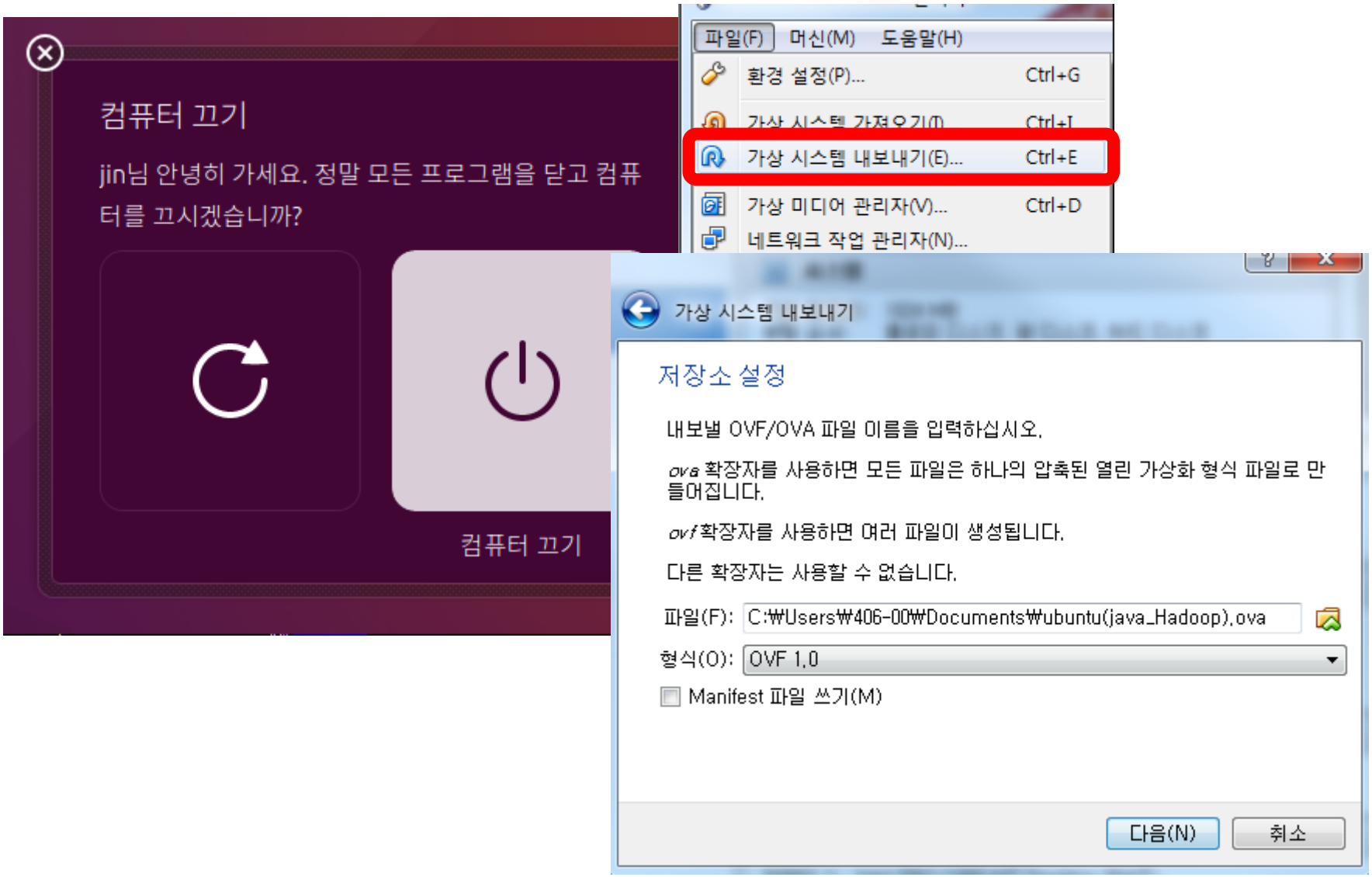
```
jin@jin-VirtualBox:~$ source /etc/profile
```

사용자 등록 및 하둡 소유권 변경

```
jin@jin-VirtualBox:~$ sudo addgroup hadoop
그룹 `hadoop' (GID 1001) 추가 ...
완료.
jin@jin-VirtualBox:~$ sudo adduser --ingroup hadoop manager
'manager' 사용자를 추가 중...
새 사용자 'manager' (1001) 을(를) 그룹 'hadoop' (으)로 추가 ...
'/home/manager' 홈 디렉터리를 생성하는 중...
'/etc/skel'에서 파일들을 복사하는 중...
새 UNIX 암호 입력:
새 UNIX 암호 재입력:
```

```
jin@jin-VirtualBox:~$ groups manager
manager : hadoop
jin@jin-VirtualBox:~$ sudo adduser manager sudo
[sudo] password for jin:
'manager' 사용자를 'sudo' 그룹에 추가 중...
사용자 manager을(를) sudo 그룹에 등록 중
완료.
```

복제하기


컴퓨터 끄기
jin님 안녕히 가세요. 정말 모든 프로그램을 닫고 컴퓨터를 끄시겠습니까?
컴퓨터 끄기
파일(F)
머신(M)
도움말(H)
환경 설정(P)... Ctrl+G
가상 시스템 내보내기(E)... Ctrl+E
가상 미디어 관리자(V)... Ctrl+D
네트워크 작업 관리자(N)...
가상 시스템 내보내기
저장소 설정
내보낼 OVF/OVA 파일 이름을 입력하십시오.
ova 확장자를 사용하면 모든 파일은 하나의 압축된 열린 가상화 형식 파일로 만들어집니다.
ovf 확장자를 사용하면 여러 파일이 생성됩니다.
다른 확장자는 사용할 수 없습니다.
파일(F): C:₩Users₩406-00₩Documents₩ubuntu(java_Hadoop).ova
형식(O): OVF 1.0
Manifest 파일 쓰기(M)
다음(N)
취소

프롬프트 변경

```
jin@jin-VirtualBox:~$ sudo vim /etc/hostname
[sudo] jin의  암호 :
jin@jin-VirtualBox:~$ sudo /bin/hostname -F /etc/hostname
jin@jin-VirtualBox:~$ sudo reboot
```

Hosts 파일 변경

```
manager@jin-VirtualBox:~$ sudo vim /etc/hosts
[sudo] manager의  암 호 :
```

```
127.0.0.1       localhost
127.0.1.1       jin-VirtualBox

192.168.56.101  master
192.168.56.102  slave1
192.168.56.103  slave2
192.168.56.104  slave3

# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
```

Static IP 설정

```
jin@jin-VirtualBox:~$ sudo vim /etc/network/interfaces
```

```
# interfaces(5) file used by ifup(8) and ifdown(8)
auto lo
iface lo inet loopback

auto enp0s8
iface enp0s8 inet static
address 192.168.56.101
netmask 255.255.255.0
gateway 192.168.56.1
dns-nameservers 168.126.63.1
```

```
jin@jin-VirtualBox:~$ sudo /etc/init.d/networking restart
[ ok ] Restarting networking (via systemctl): networking.service.
```

Master 설정 2

Masters 및 slaves 파일 수정

```
$ sudo vim /usr/local/hadoop/etc/hadoop/slaves
```

```
slave1
slave2
slave3
```

hdfs-site 수정

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

```
<configuration>
        <property>
                        <name>dfs.replication</name>
                        <value>3</value>
        </property>
        <property>
                        <name>dfs.namenode.name.dir</name>
                        <value>file:/usr/local/hadoop/hdfs/namenode</value>
        </property>
        <property>
                        <name>dfs.http.address</name>
                        <value>master:50070</value>
        </property>
        <property>
                        <name>dfs.secondary.http.address</name>
                        <value>slave1:50090</value>
        </property>
```

Slave에 인증키 배포

```
manager@master:~$ ssh-copy-id -i  .ssh/id_rsa.pub manager@slave1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: ".ssh/id_rsa.pub"
The authenticity of host 'slave1 (192.168.56.102)' can't be established.
ECDSA key fingerprint is SHA256:CR2ZcVni0W+B8hD3yBIr5npVLucsoF9NBmWEja+qHvI.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter
lled
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompt
w keys
manager@slave1's password:

Number of key(s) added: 1

Now try logging into the machine, with:   "ssh 'manager@slave1'"
and check to make sure that only the key(s) you wanted were added.
```

배포하기

Master & slave

master

```
$ scp hadoop.tar manager@slave1:/home/manager

$ ssh manager@slave1 "cd /home/manager;
                      tar xf hadoop.tar;rm hadoop.tar"
```

slave

```
$ sudo mkdir /usr/local/hadoop

$ cd usr/local/hadoop/
/usr/local/hadoop$ sudo mv * /usr/local/hadoop/

$ sudo chown -R manager:hadoop /usr/local/hadoop/
```

Slave 설정 2

~/.bashrc 수정

```
$   sudo vim .bashrc
```

```
#HADOOP VARIABLES START
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
```

```
$ sudo mkdir -p /usr/local/hadoop/hdfs/datanode
```

```
$ sudo chown -R manager:hadoop /usr/local/hadoop/hdfs
```

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

```
<property>
        <name>dfs.datanode.data.dir</name>
        <value>file:/usr/local/hadoop/hdfs/datanode</value>
</property>
```

실행

실행

```
$ hadoop namenode -format
$ start-all.sh
```

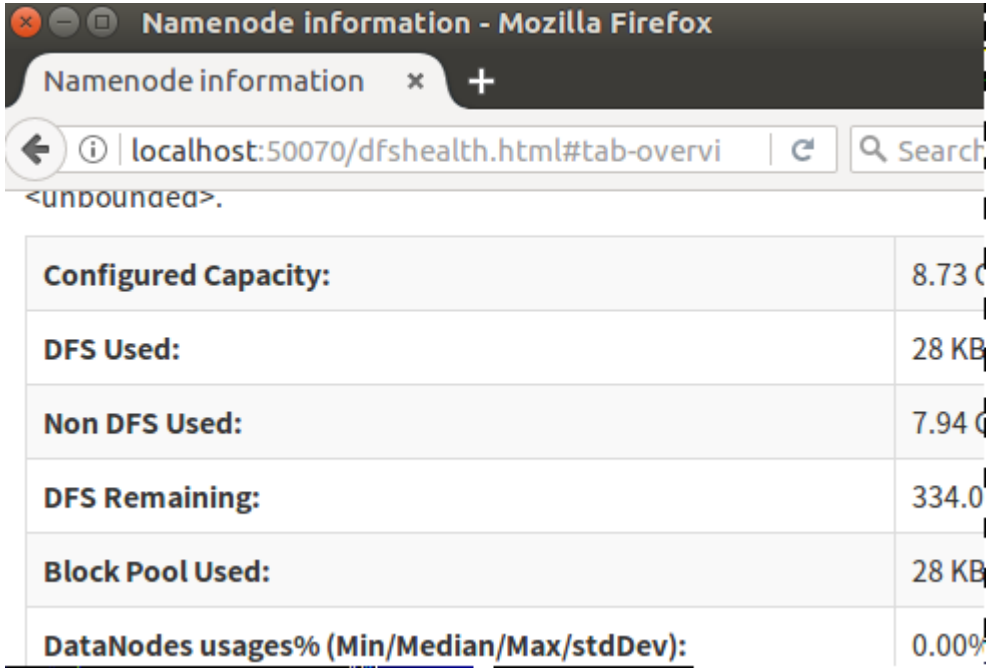
```
manager@master:~$ jps
10439 ResourceManager
10510 Jps
10158 NameNode
```

```
manager@slave2:~$ jps
4275 NodeManager
4170 DataNode
4349 Jps
```

```
manager@slave1:~$ jps
5585 DataNode
5685 SecondaryNameNode
5868 Jps
5774 NodeManager
```

```
manager@slave3:~$ jps
4624 NodeManager
4519 DataNode
4717 Jps
```

웹 브라우저 접속 및 확인


Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-overvi
Search
<unbounded>.
Configured Capacity:
8.73
DFS Used:
28 KB
Non DFS Used:
7.94
DFS Remaining:
334.0
Block Pool Used:
28 KB
DataNodes usages% (Min/Median/Max/stdDev):
0.00%