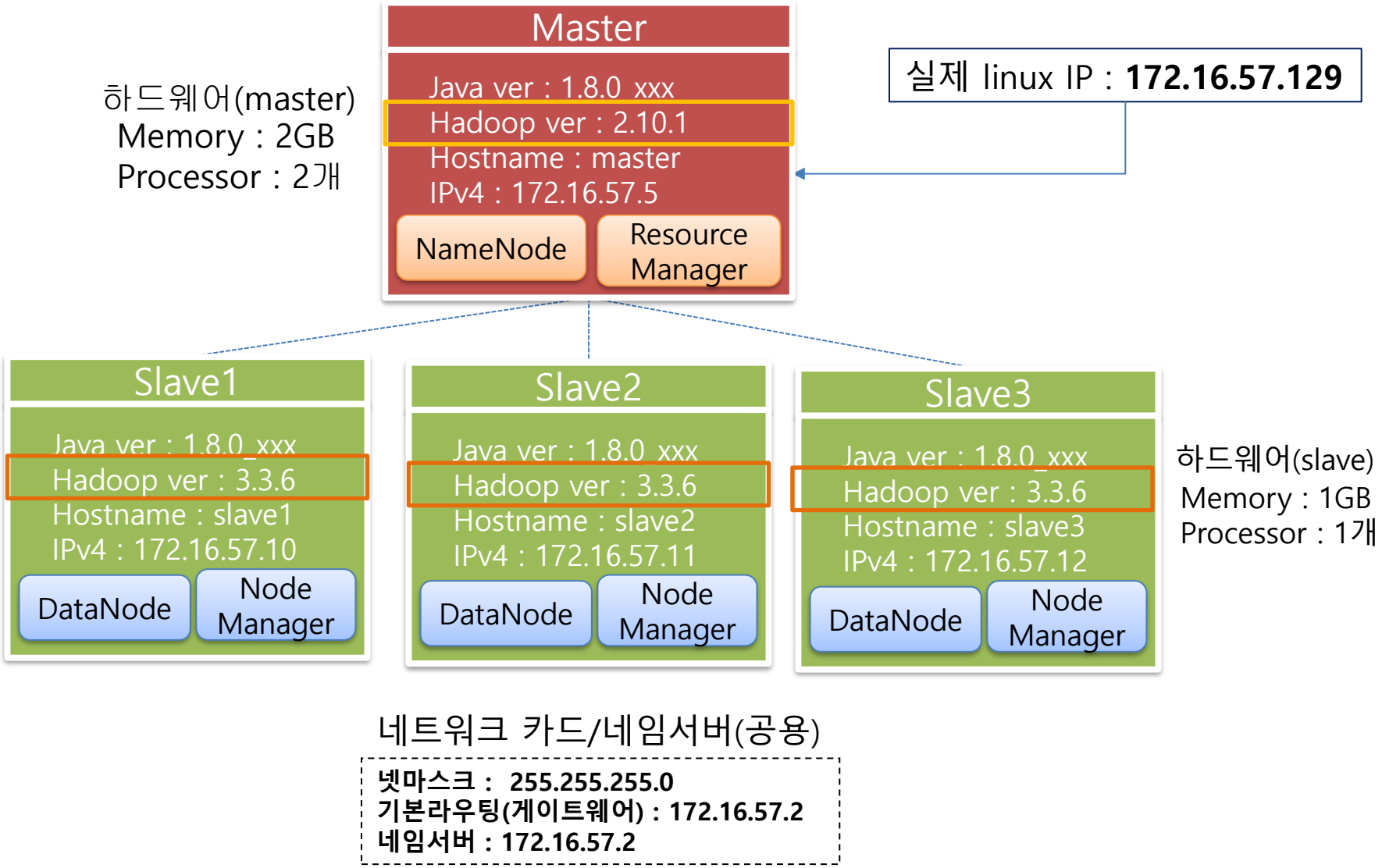
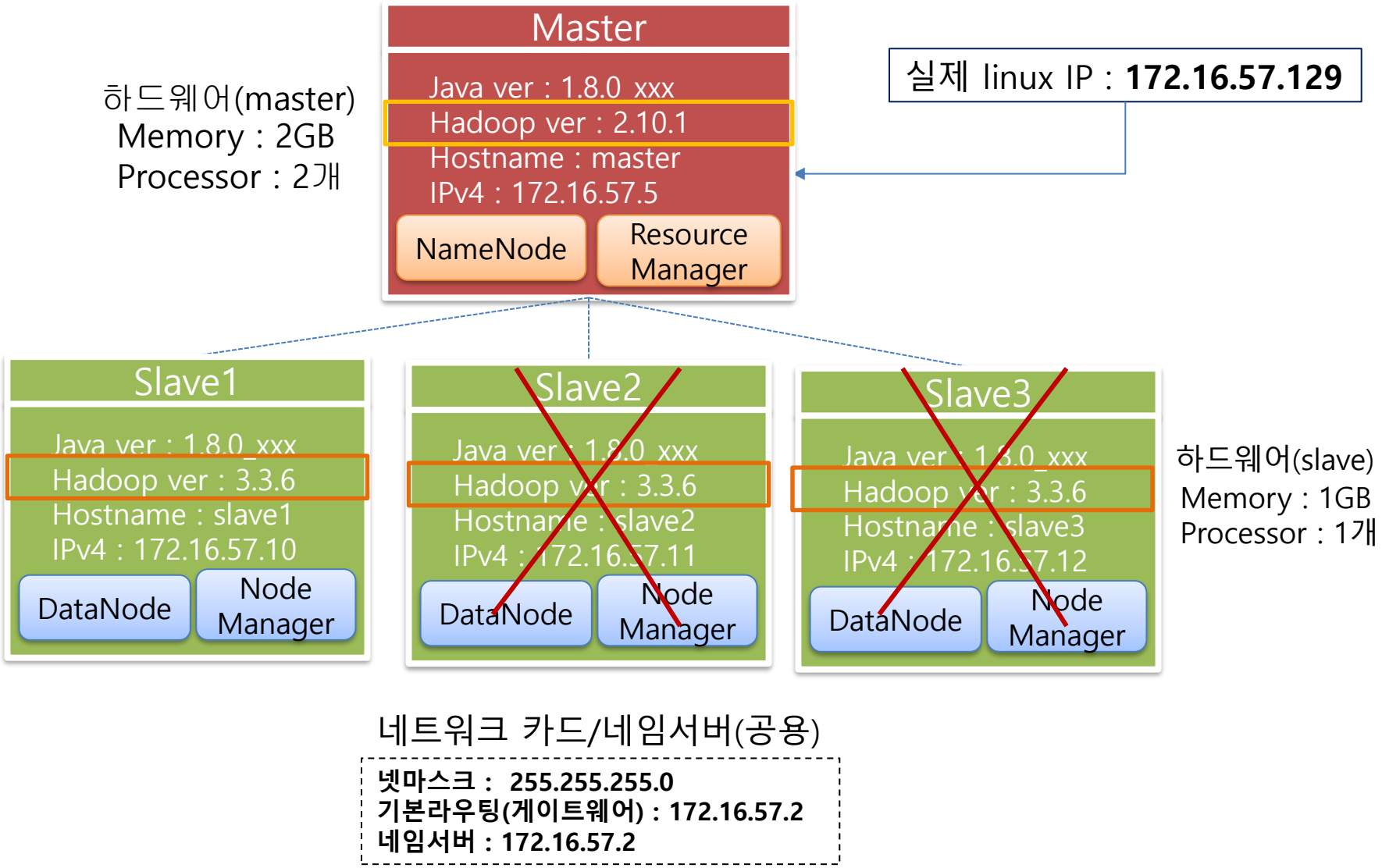
4. Hadoop 설치

목 차

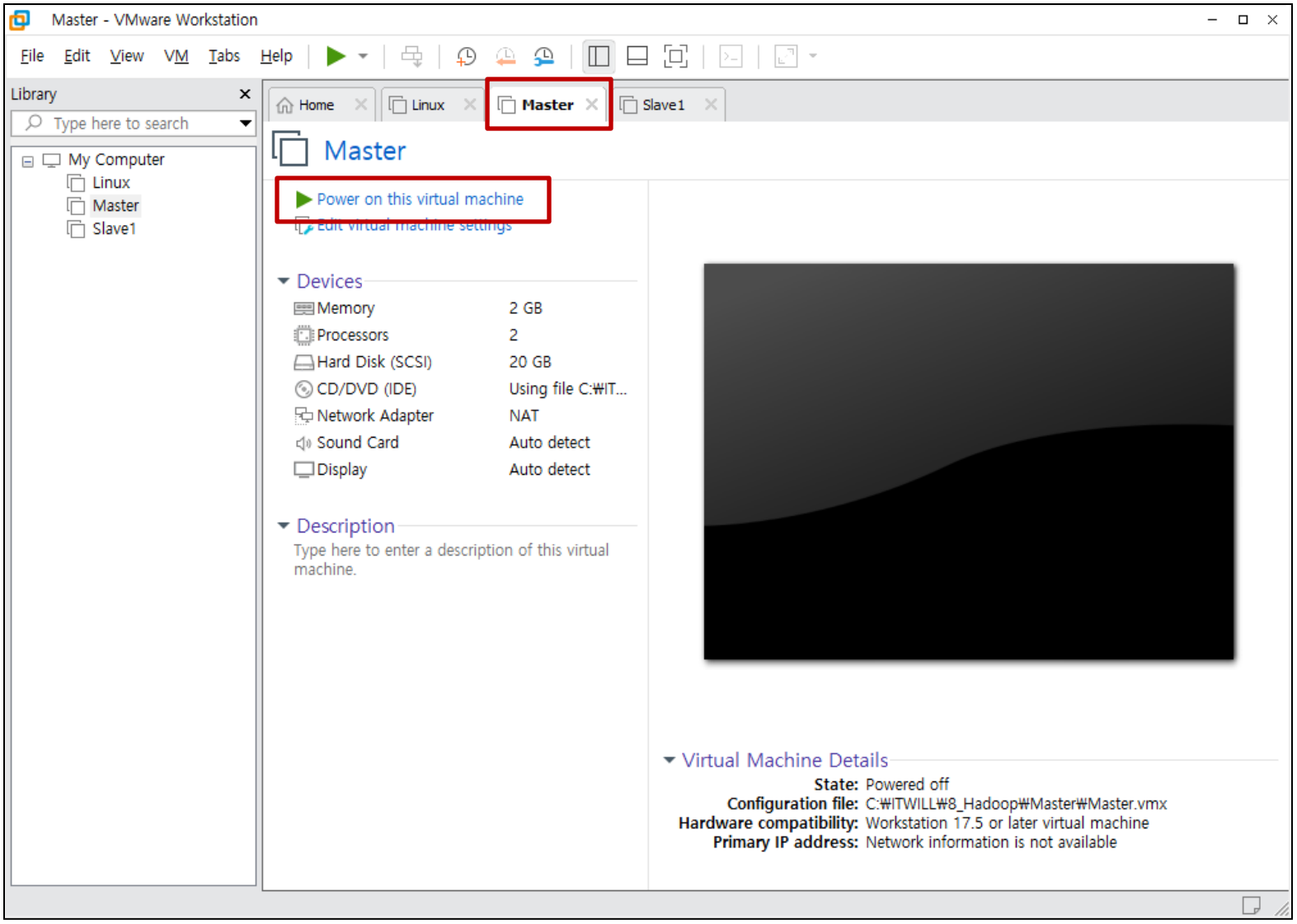
Hadoop 클러스터 구축 시스템 구성도



Hadoop 클러스터 구축 시스템 구성도 (실제 수업환경 : 속도/용량 고려)



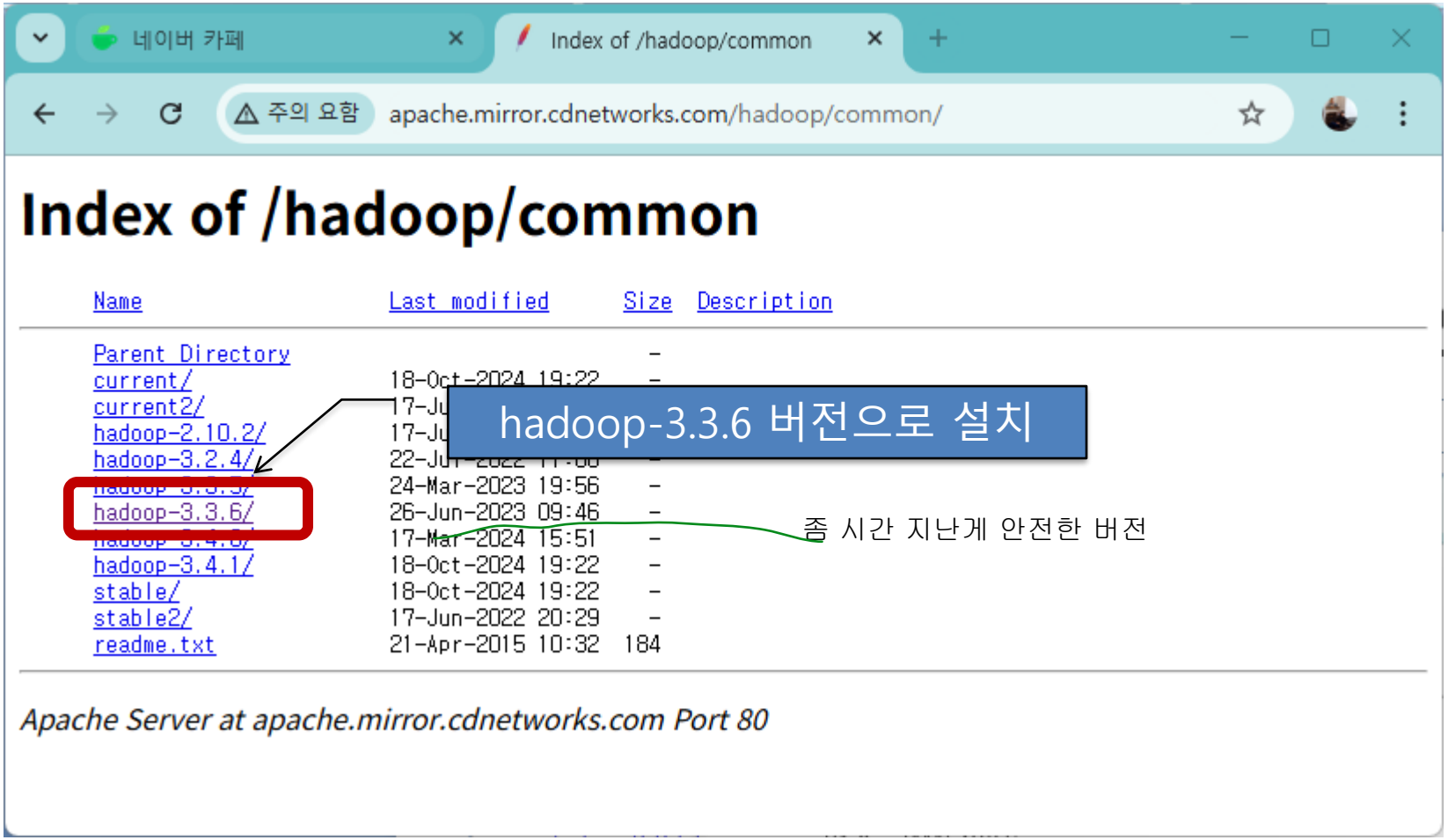
Master 서버 수행


Master - VMware Workstation
File Edit View VM Tabs Help
Library
Type here to search
My Computer
Linux
Master
Slave1
Home
Linux
Master
Slave1
Master
Power on this virtual machine
Devices
Memory 2 GB
Processors 2
Hard Disk (SCSI) 20 GB
CD/DVD (IDE) Using file C:₩IT...
Network Adapter NAT
Sound Card Auto detect
Display Auto detect
Description
Type here to enter a description of this virtual machine.
Virtual Machine Details
State: Powered off
Configuration file: C:₩ITWILL₩8_Hadoop₩Master₩Master.vmx
Hardware compatibility: Workstation 17.5 or later virtual machine
Primary IP address: Network information is not available

이제 진짜 hadoop 설치

Hadoop 설치 가능 버전 확인

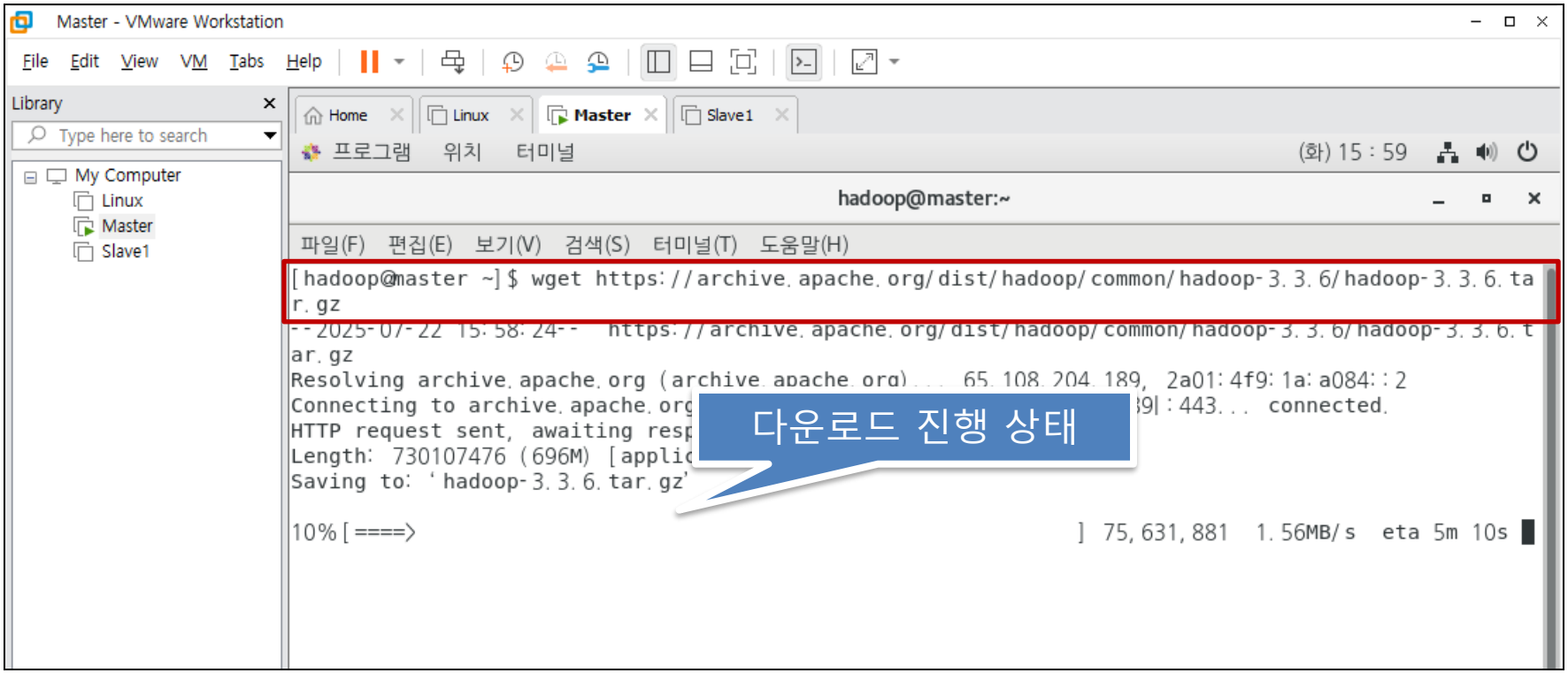
http://apache.mirror.cdnetworks.com/hadoop/common/ 에서 다운로드 가능한 버전 확인



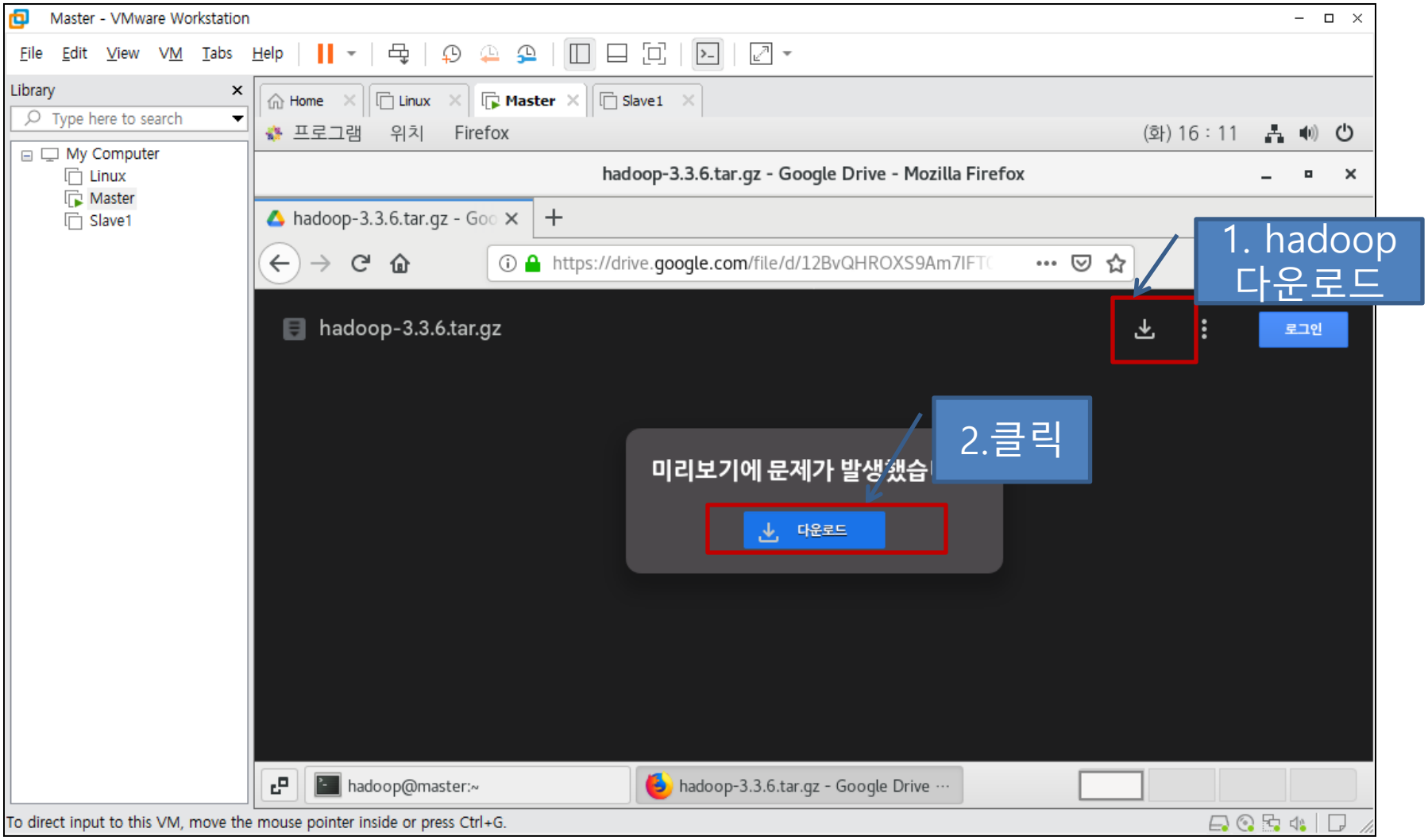
1. Hadoop 다운로드 & 설치

단계1 : Hadoop 다운로드(Master 작업)

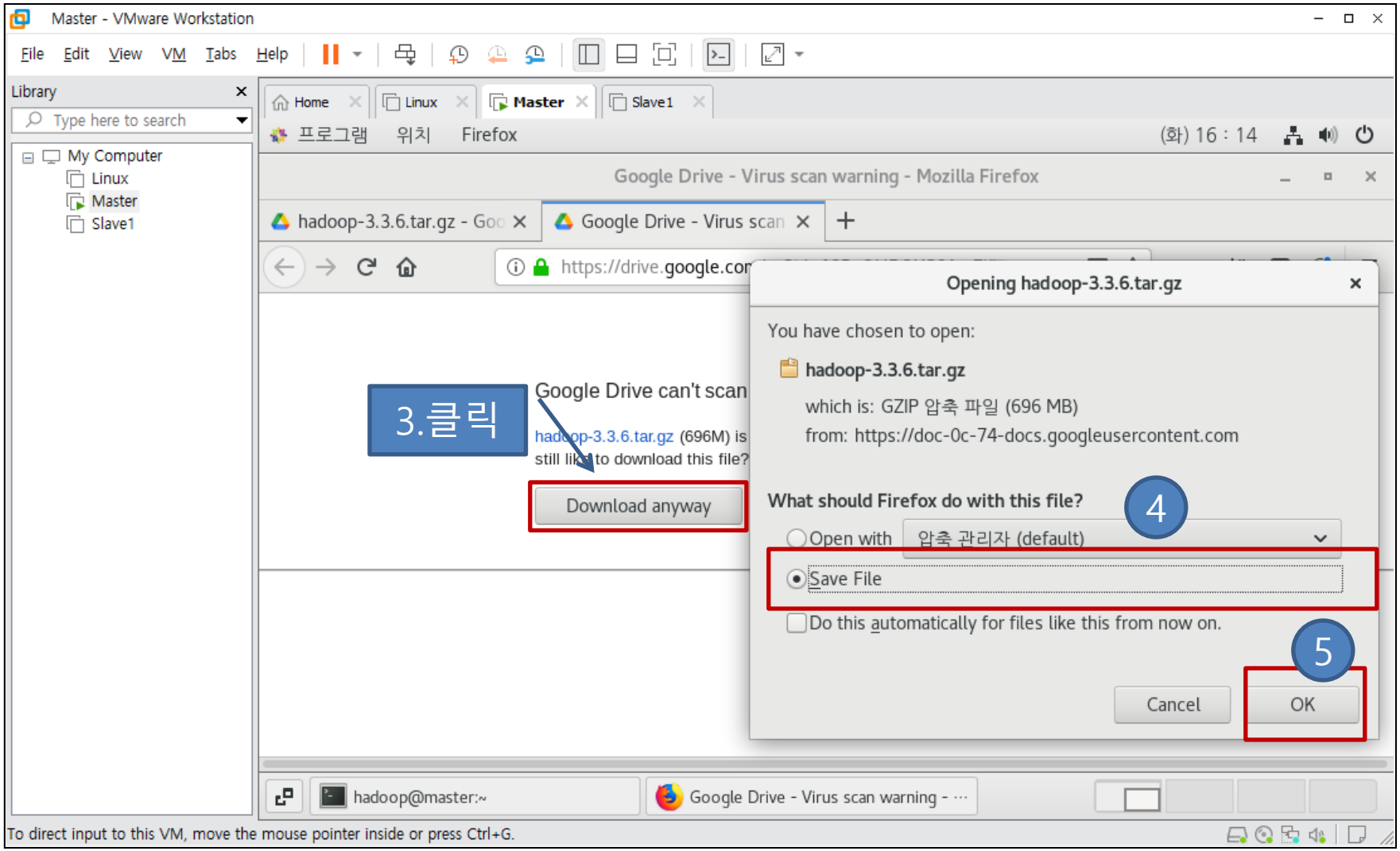
[방법1] wget 명령어로 다운받기 형식) $ wget URL



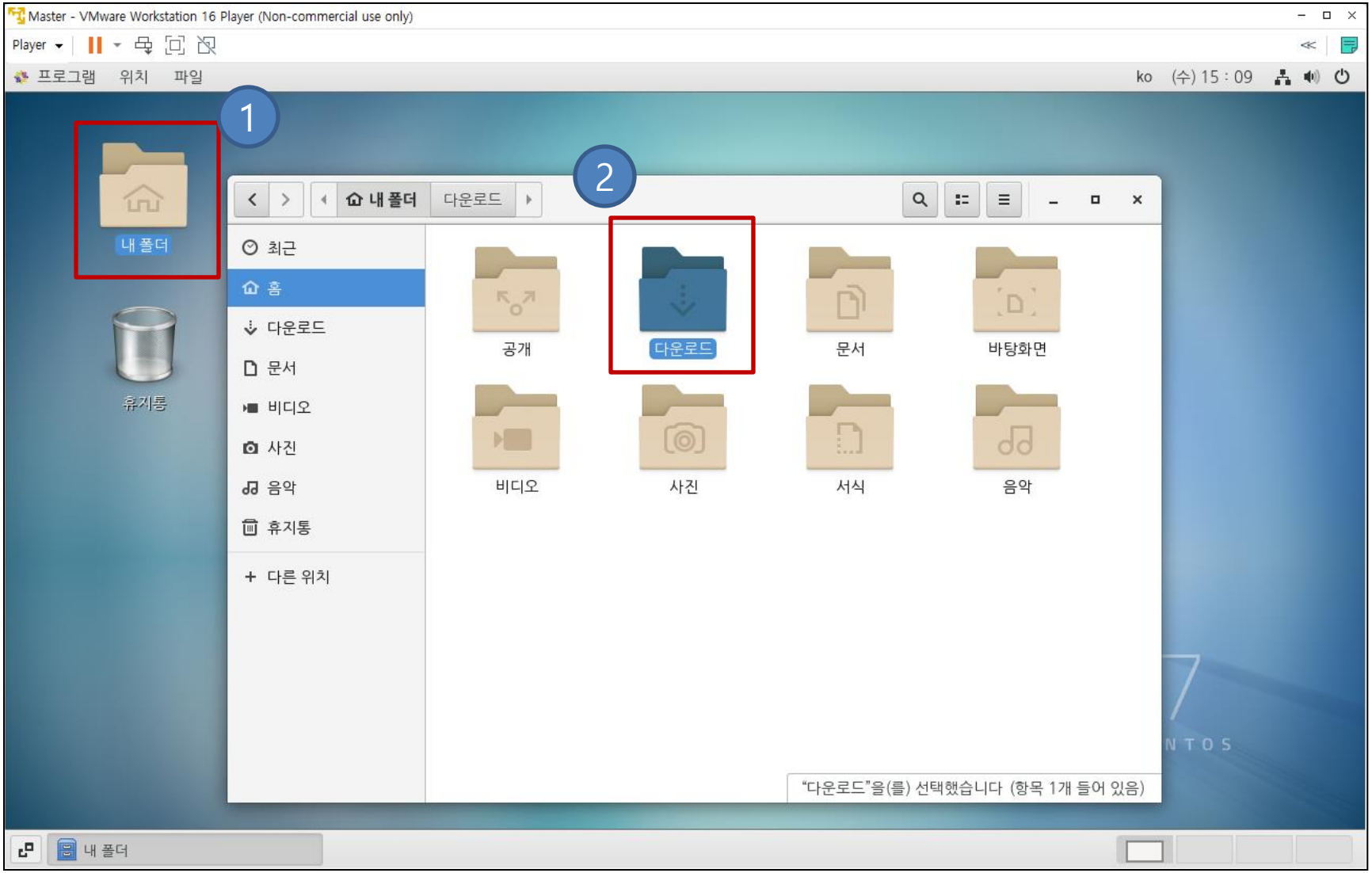
[방법2] 가상머신의 Firefox에서 Google 드라이버의 링크 주소 접속 : https://drive.google.com/file/d/12BvQHROXS9Am7IFT0DR4VKXswgFU5Y4m/view?usp=sharing



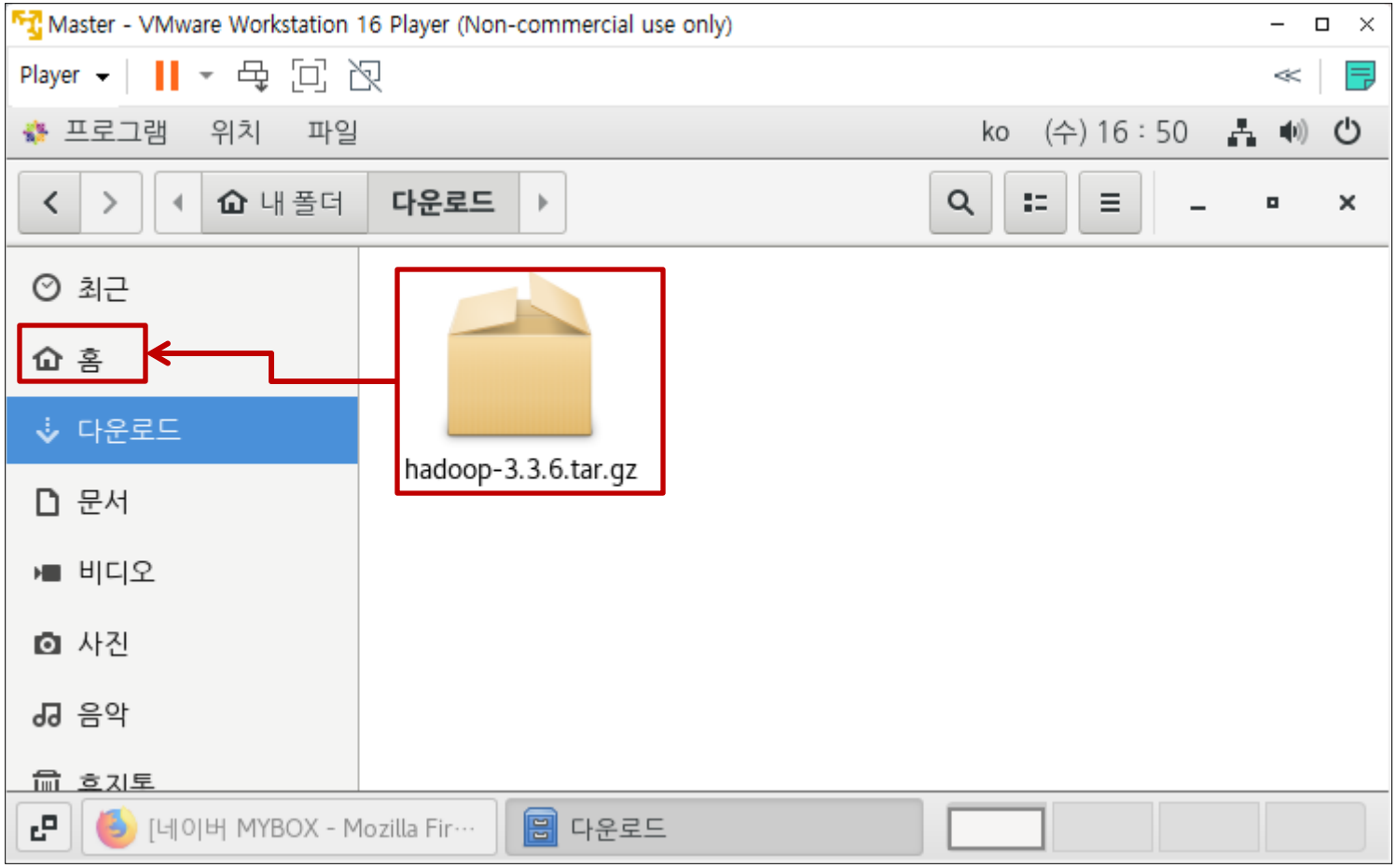
● 저장 위치 선택


Master - VMware Workstation
File Edit View VM Tabs Help
Library
Type here to search
My Computer
Linux
Master
Slave1
Home
Linux
Master
Slave1
프로그램 위치 Firefox
(화) 16 : 14
Google Drive - Virus scan warning - Mozilla Firefox
hadoop-3.3.6.tar.gz - Goo
Google Drive - Virus scan
https://drive.google.com
3.클릭
Google Drive can't scan
hadoop-3.3.6.tar.gz (696M) is
still like to download this file?
Download anyway
Opening hadoop-3.3.6.tar.gz
You have chosen to open:
hadoop-3.3.6.tar.gz
which is: GZIP 압축 파일 (696 MB)
from: https://doc-0c-74-docs.googleusercontent.com
What should Firefox do with this file?
4
Open with
압축 관리자 (default)
Save File
Do this automatically for files like this from now on.
5
Cancel
OK
hadoop@master:~
Google Drive - Virus scan warning - ···
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.

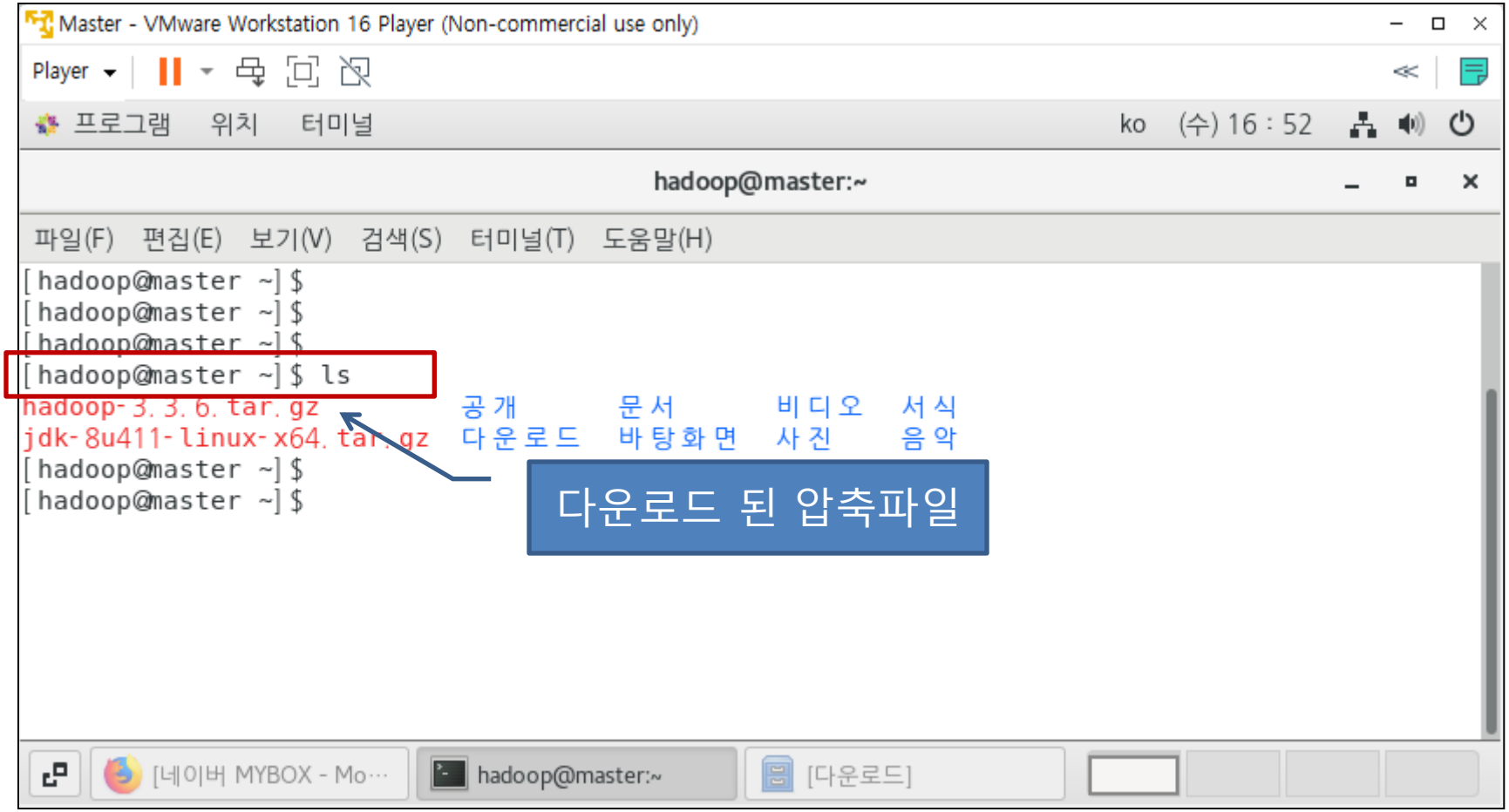
● 다운로드 디렉터리 선택


Master - VMware Workstation 16 Player (Non-commercial use only)
Player
프로그램 위치 파일
ko (수) 15 : 09
1
2
내 폴더
휴지통
내 폴더 다운로드
최근
홈
다운로드
문서
비디오
사진
음악
휴지통
다른 위치
공개
다운로드
문서
바탕화면
비디오
사진
서식
음악
"다운로드"을(를) 선택했습니다 (항목 1개 들어 있음)
내 폴더

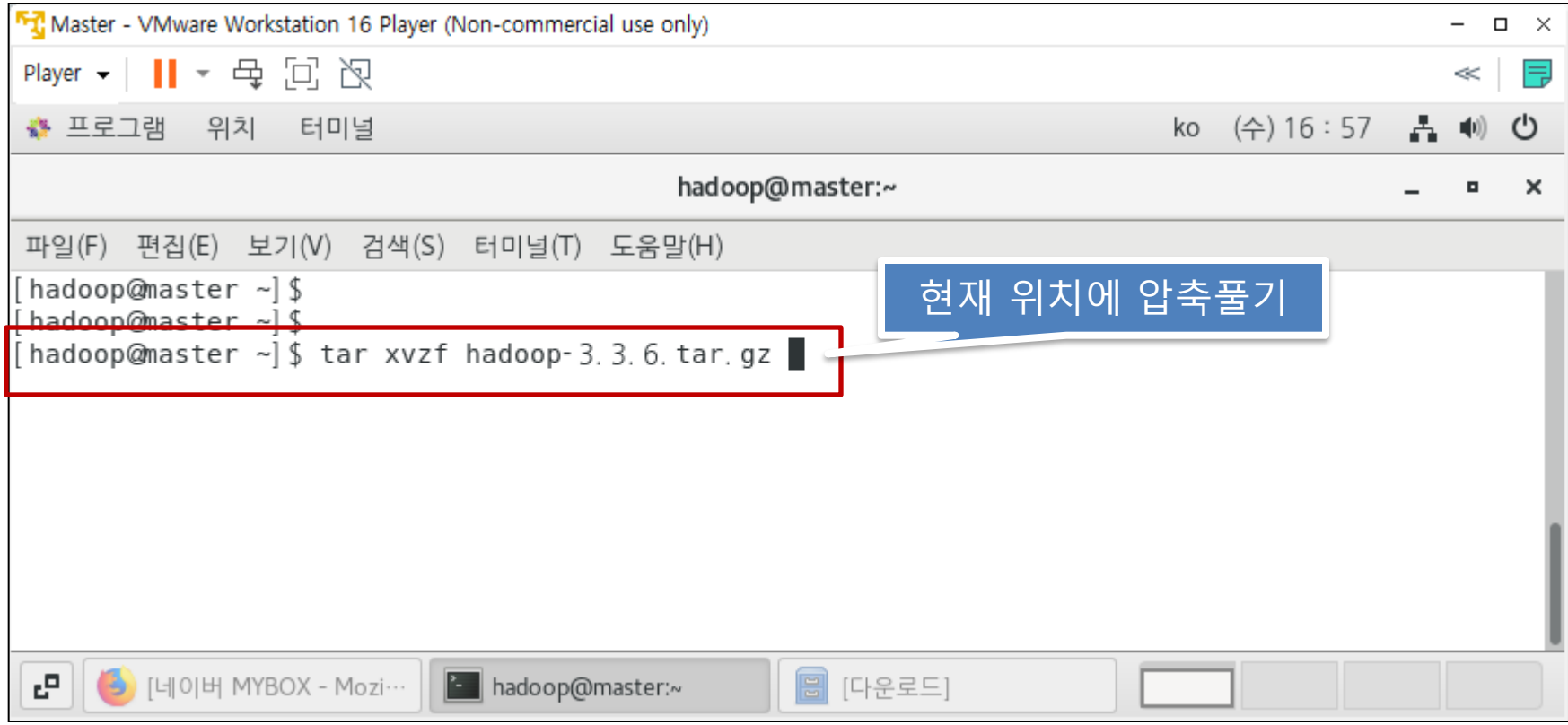
- Hadoop 압축파일을 [홈] 으로 끌어다 & 놓기



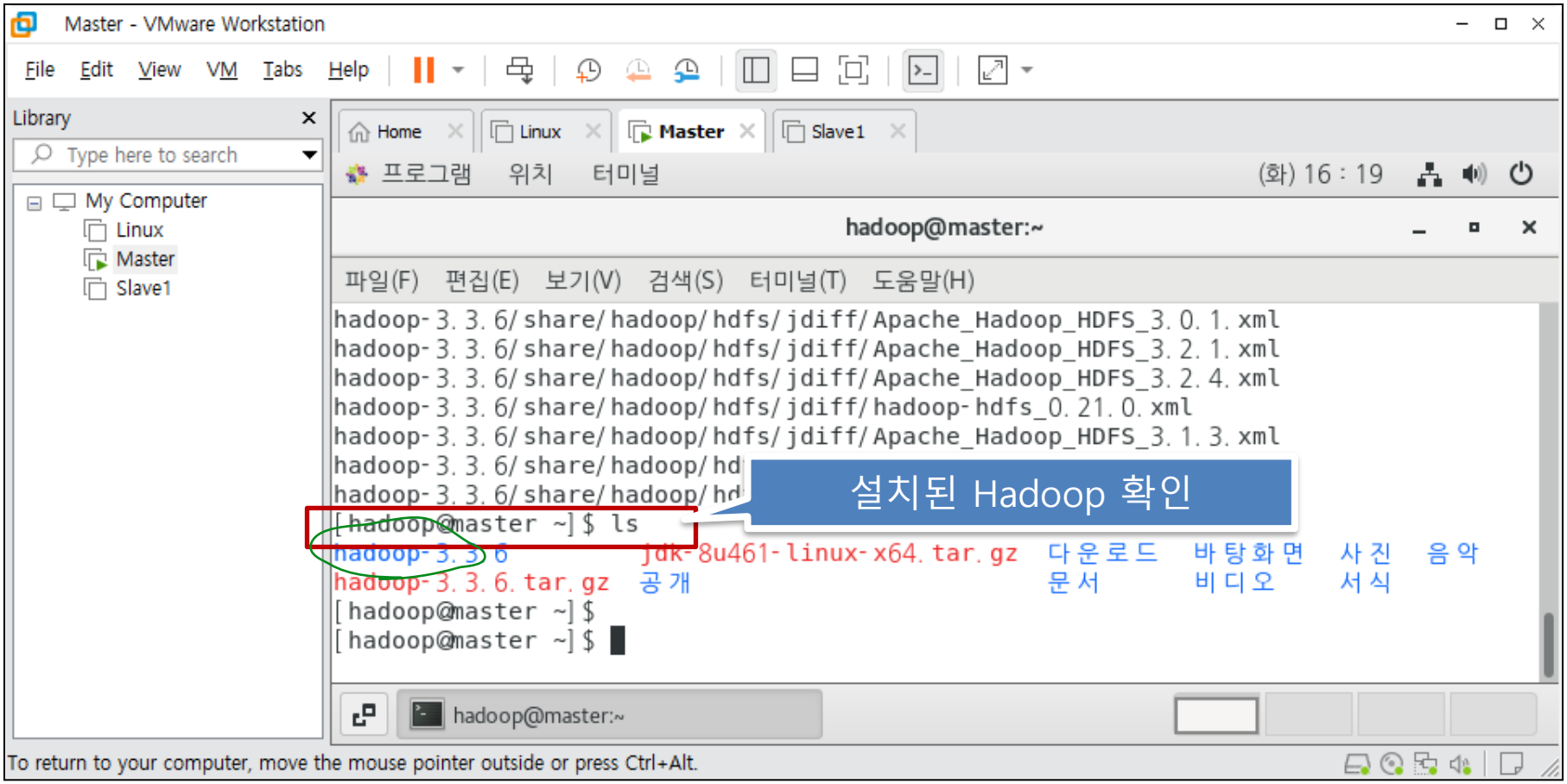
- hadoop [홈 디렉터리]에서 다운로드 파일 확인



단계2 : Hadoop 설치(압축풀기)

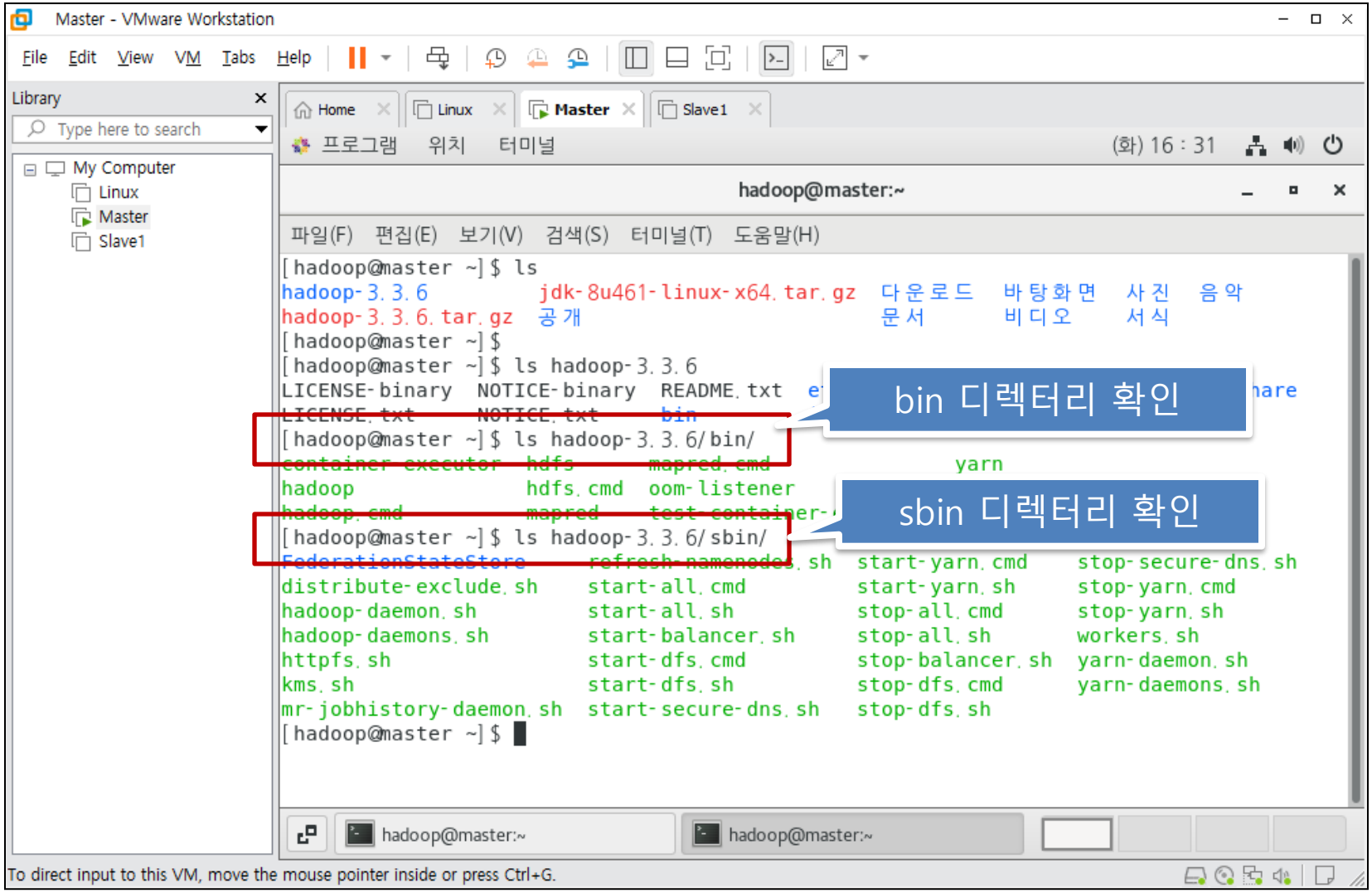


단계3 : Hadoop 설치 확인



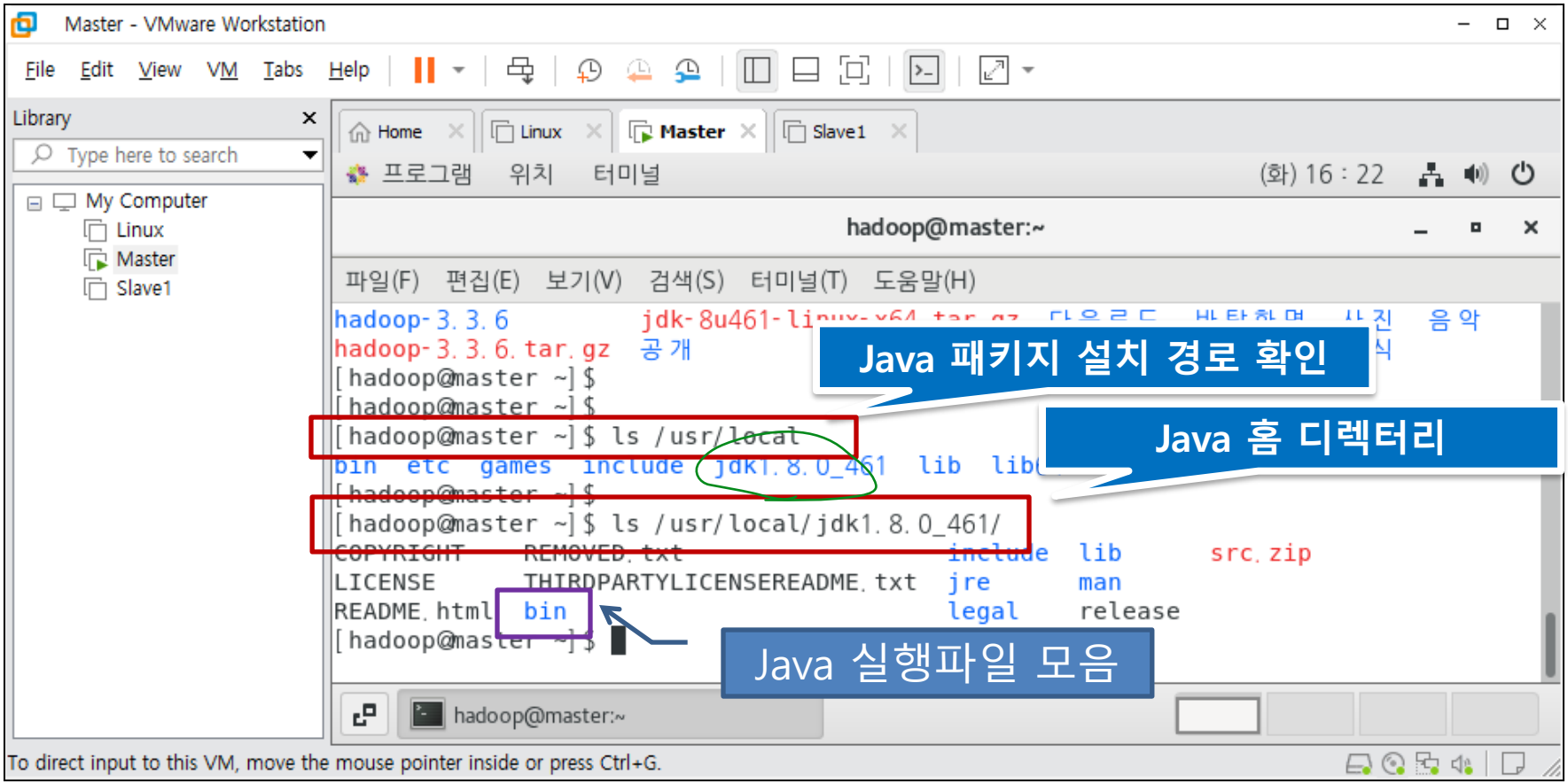
❖ 현재 설치된 Hadoop 버전 : hadoop-3.3.6

단계4 : Hadoop 실행 파일 확인

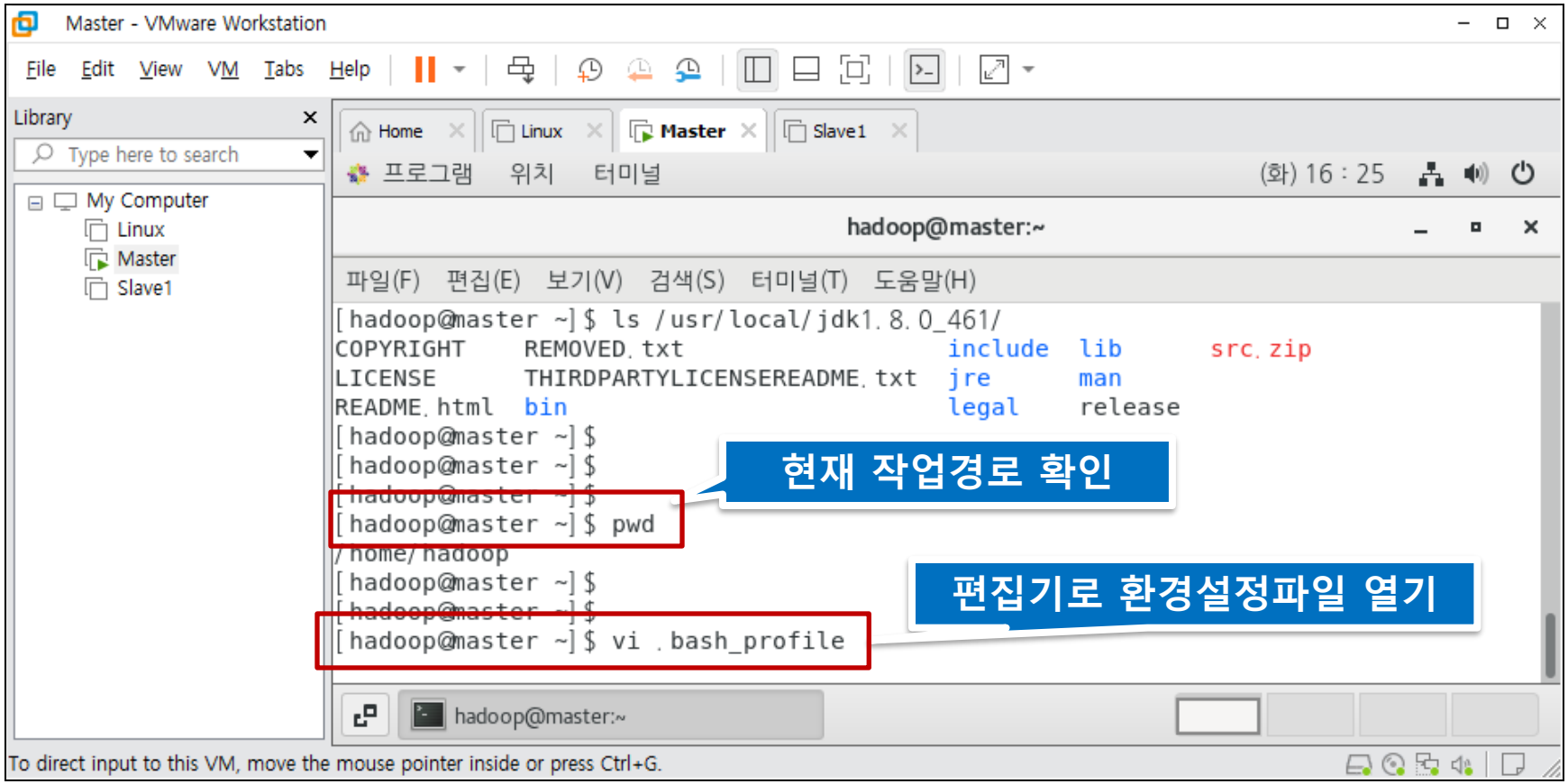


2. 환경설정 : Java/Hadoop 홈 Path 설정

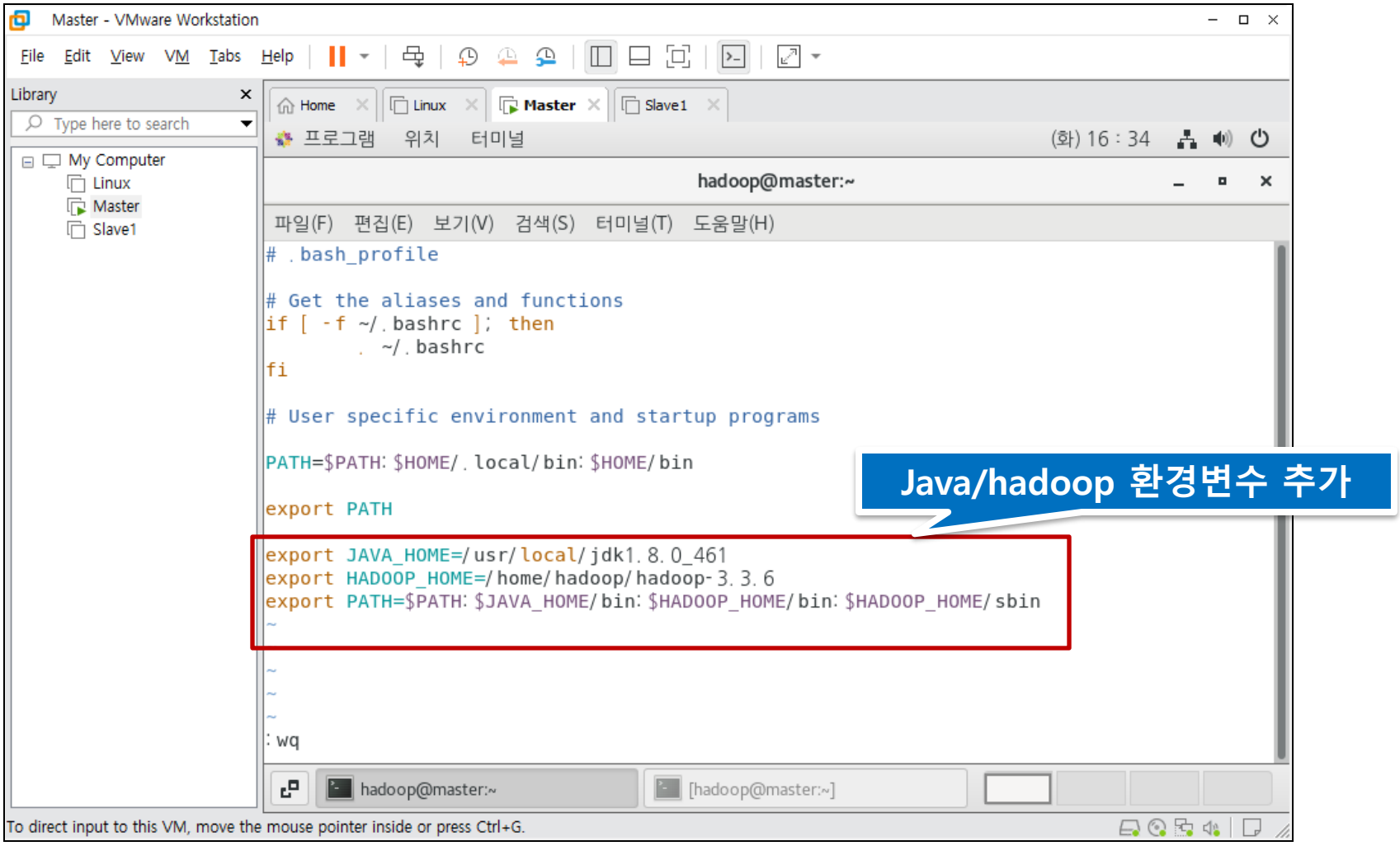
1) Java 홈 디렉터리 확인 : 환경설정 전에 홈 디렉터리 확인



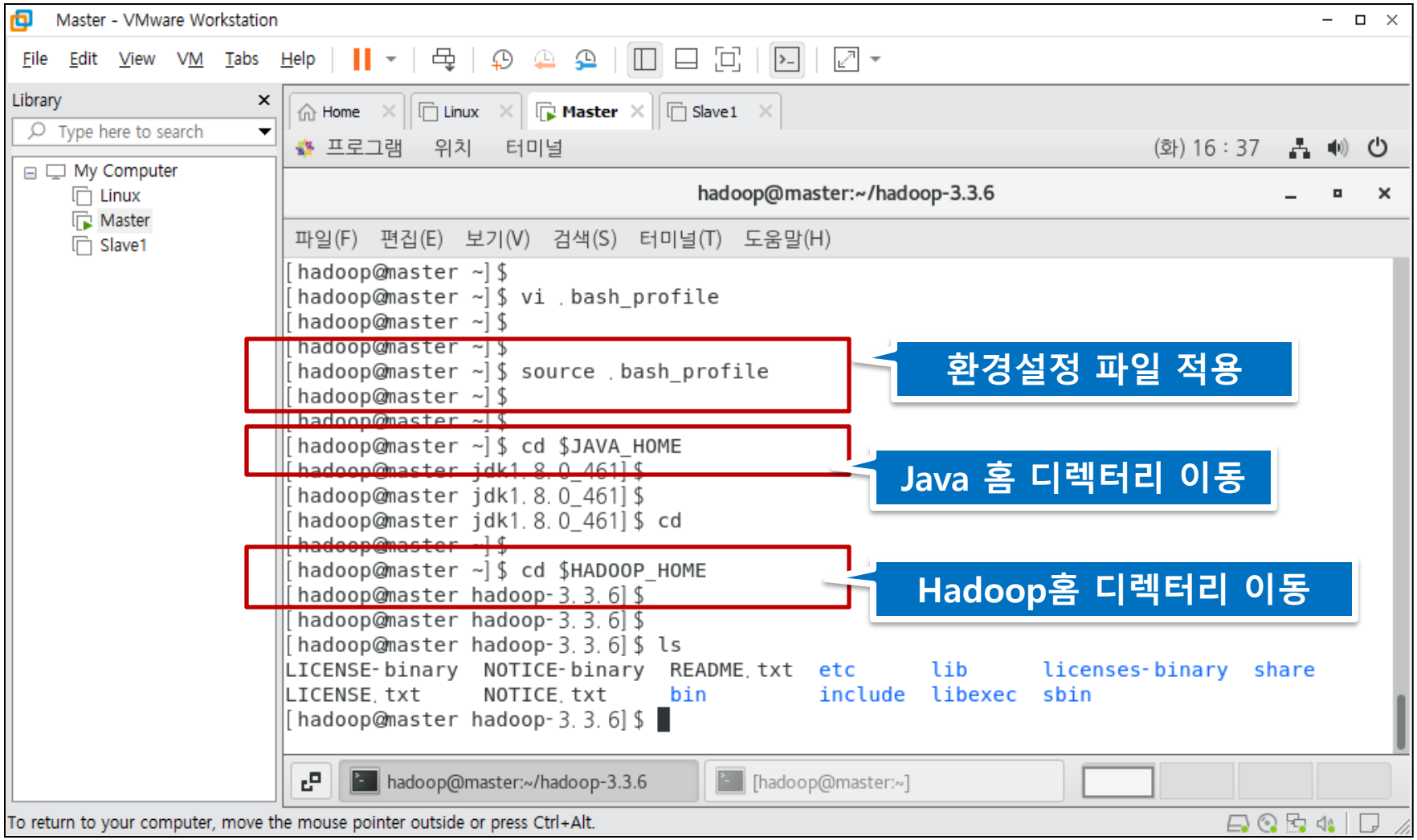
2) 사용자 환경설정 파일(.bash_profile)에서 환경설정



3) Java & Hadoop 홈 환경변수와 PATH 환경변수 지정



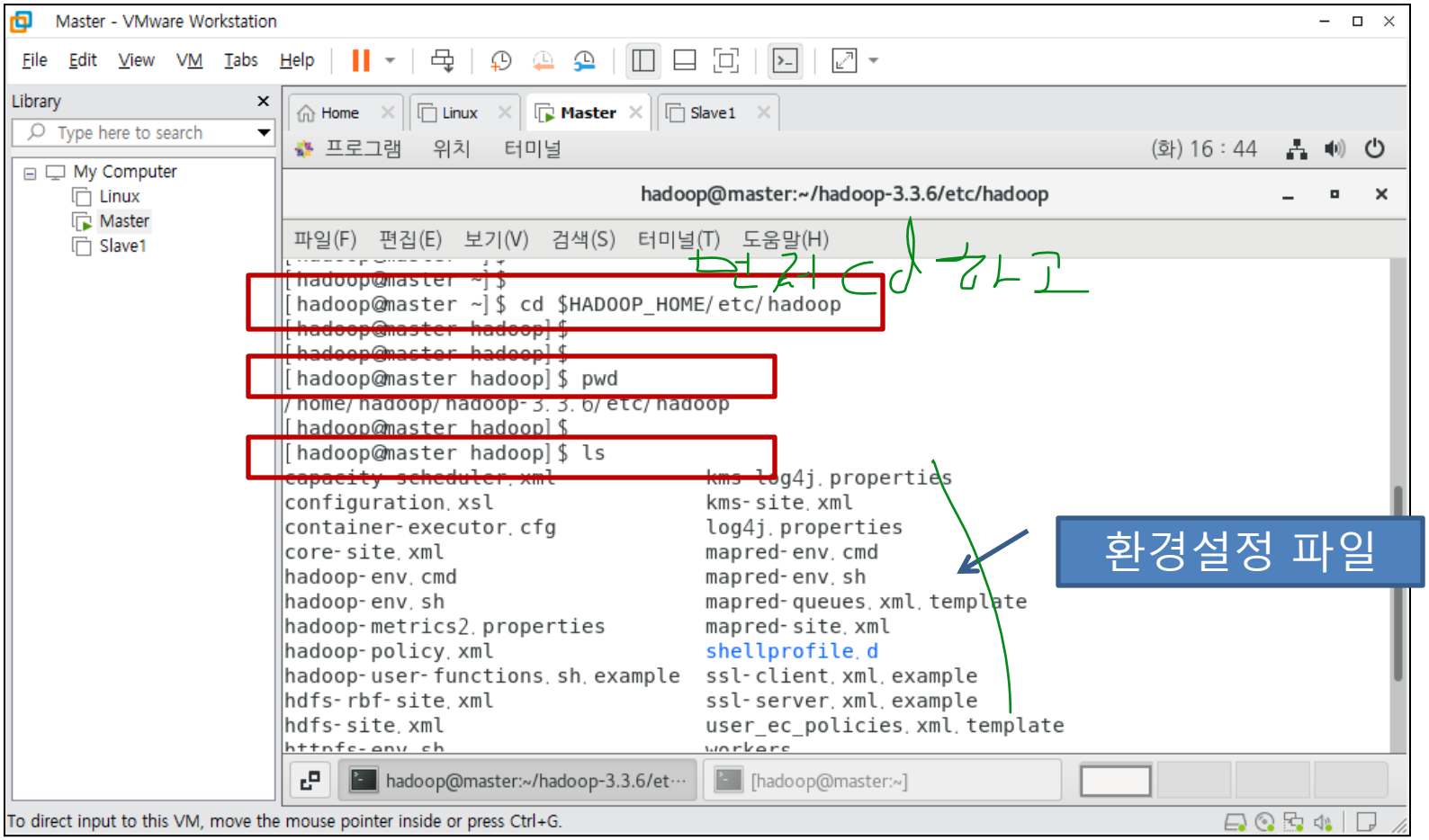
4) .bash_profile 적용/환경변수 확인



3. Hadoop 환경 설정 관련 파일

설정 파일	설정 내용
hadoop-env.sh	Hadoop을 실행하는 설정파일(JDK경로, Class 패스, 데몬 실행 등)
core-site.xml	분산파일시스템(HDFS)의 URL 지정
mapred-site.xml	Mapreduce에서 사용할 환경설정 파일
hdfs-site.xml	NameNode와 DataNode의 디렉터리 지정
yarn-site.xml	YARN(Hadoop 클러스터 관리도구) 설정파일, Resource Manager, Node Manager 지정
workers	DataNode를 실행할 서버(Slave) 설정

- Hadoop 환경설정 파일이 있는 경로 이동

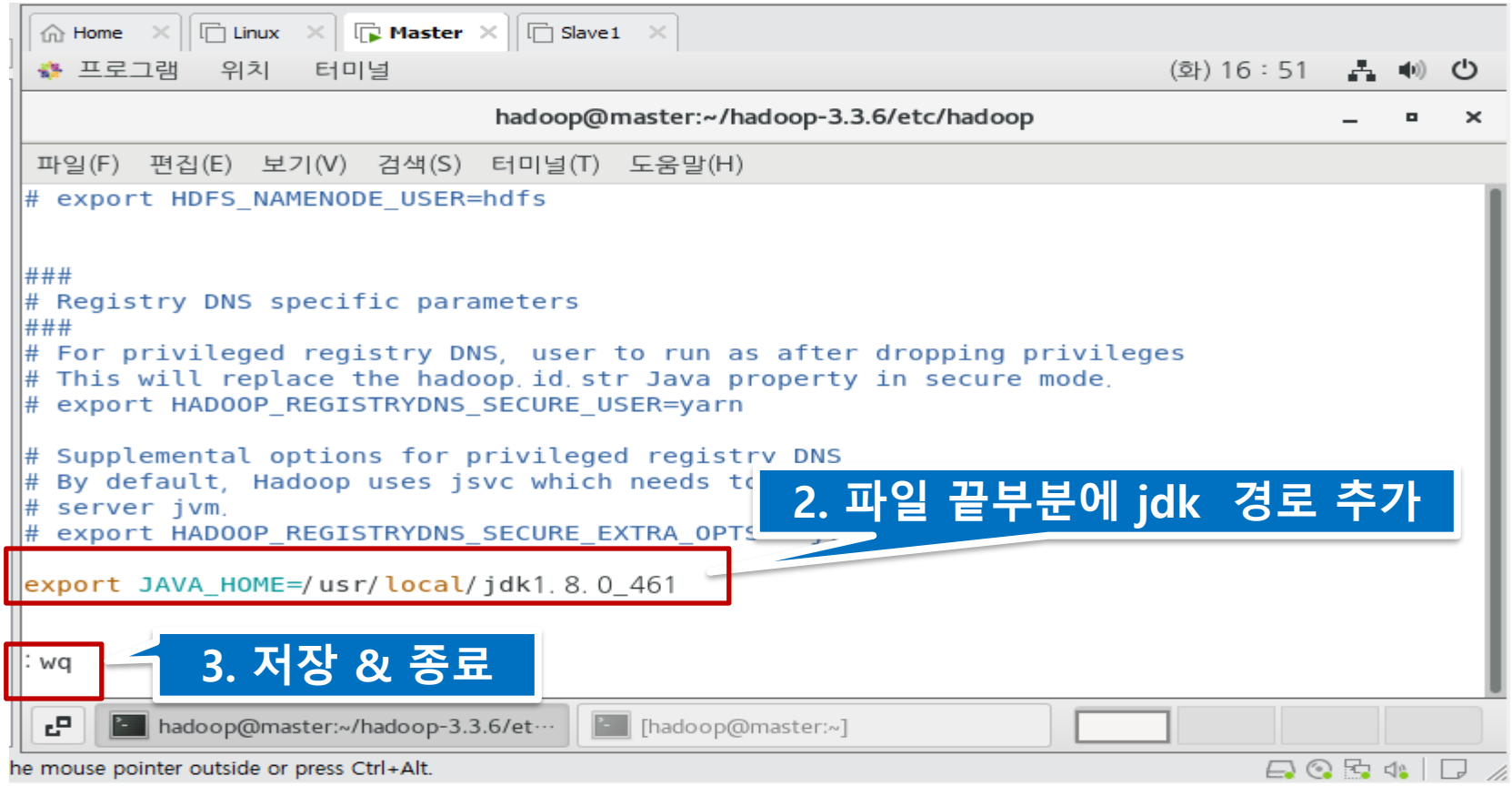


1) hadoop-env.sh(Hadoop 실행 설정파일)

Hadoop 환경 설정 (jdk 추가)

[hadoop@master hadoop]$ **vi hadoop-env.sh**

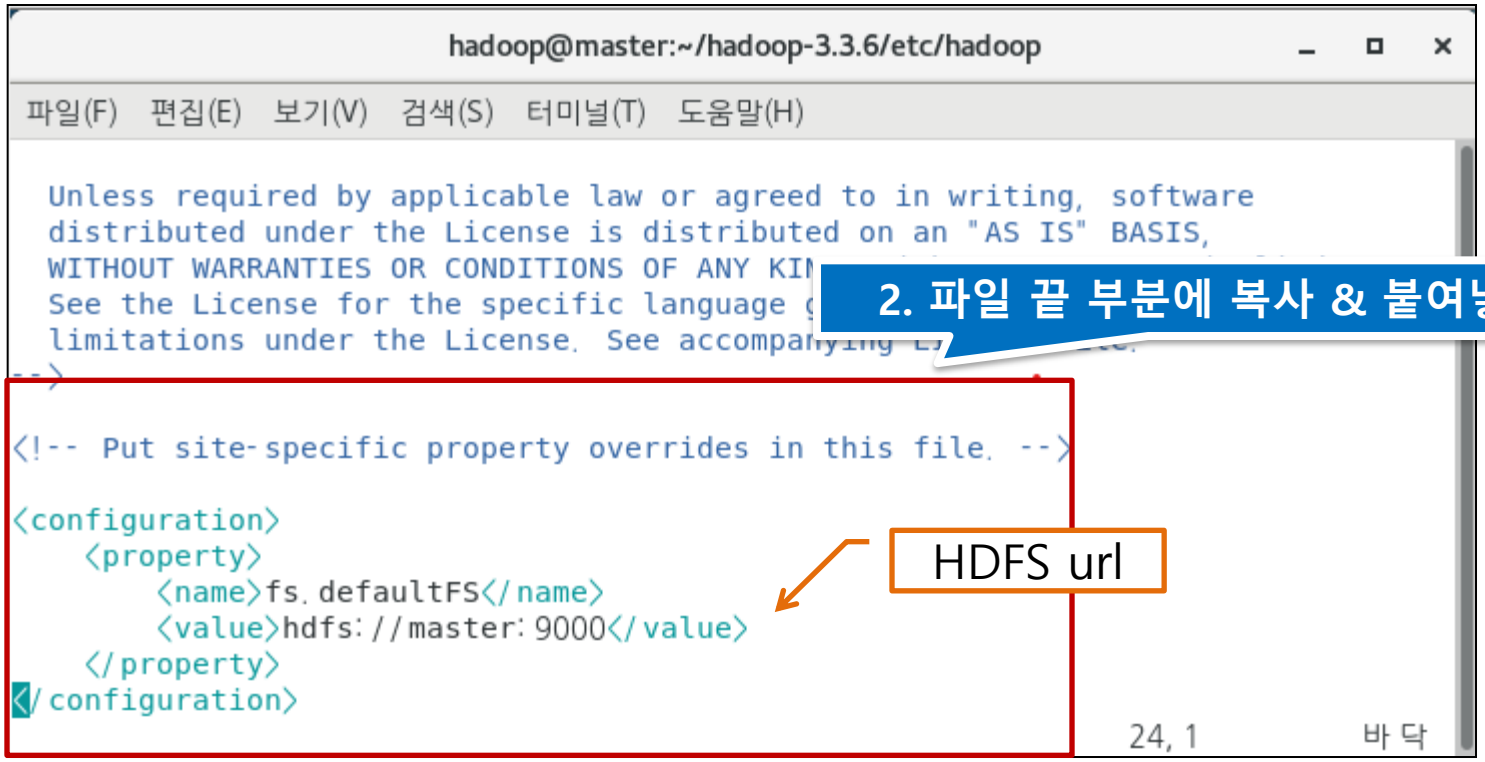
1. 설정파일 열기



2) HDFS의 NameNode URL 지정

[hadoop@master hadoop]$ **vi core-site.xml**

1. 설정파일 열기



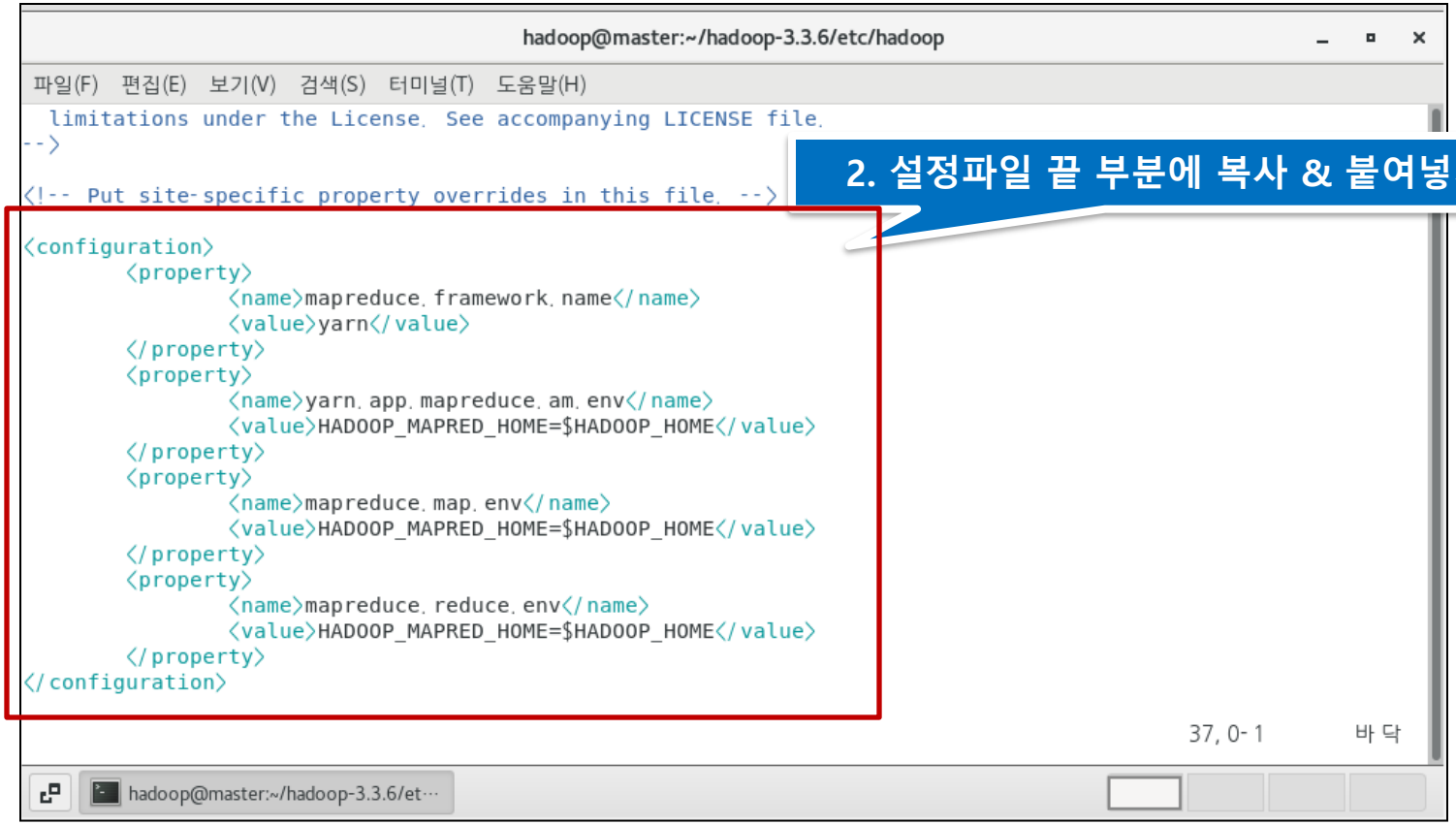
2. 파일 끝 부분에 복사 & 붙여넣기

HDFS url

3) MapReduce 관련 설정

[hadoop@master hadoop]$ **vi mapred-site.xml**

1. 설정파일 열기



2. 설정파일 끝 부분에 복사 & 붙여넣기

4) namenode와 datanode 디렉터리 생성 (hdfs-site.xml 에서 지정한 디렉터리)

Hadoop 홈 디렉토리 아래 namenode와 datanode 디렉토리 생성

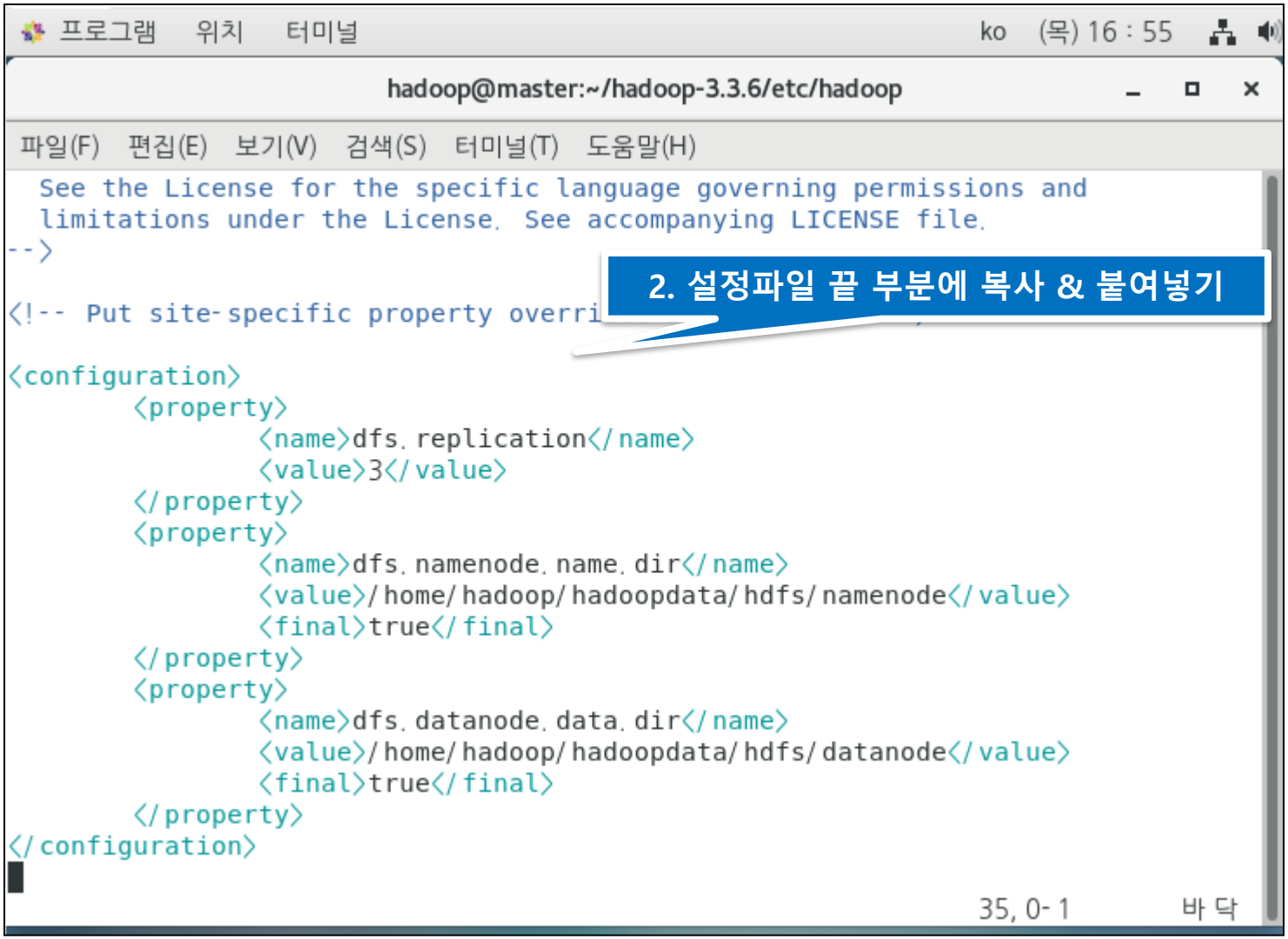
```
[hadoop@master hadoop]$ mkdir -p ~/hadoopdata/hdfs/namenode
[hadoop@master hadoop]$ mkdir -p ~/hadoopdata/hdfs/datanode
```

5) 하둡분산파일시스템(HDFS) 환경설정

Multi-Node Cluster 모드용

[hadoop@master hadoop]$ **vi hdfs-site.xml**

1. 설정파일 열기



2. 설정파일 끝 부분에 복사 & 붙여넣기

6) YARN 설정(Resource & Node Manager 지정)

Multi-Node Cluster 모드용

[hadoop@master hadoop]$ **vi yarn-site.xml**

1. 설정파일 열기

2. 설정파일 끝 부분에 복사 & 붙여넣기

hadoop@master:~/hadoop-3.3.6/etc/hadoop

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
<configuration>

<!-- Site specific YARN configuration properties -->
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>master</value>
        </property>
        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
        <property>
                <name>yarn.nodemanager.vmem-check-enabled</name>
                <value>false</value>
        </property>
        <property>
                <name>yarn.resourcemanager.address</name>
                <value>master:8032</value>
        </property>
        <property>
                <name>yarn.resourcemanager.scheduler.address</name>
                <value>master:8030</value>
        </property>
        <property>
                <name>yarn.resourcemanager.resource-tracker.address</name>
                <value>master:8031</value>
        </property>

</configuration>
```

45,0-1 바닥

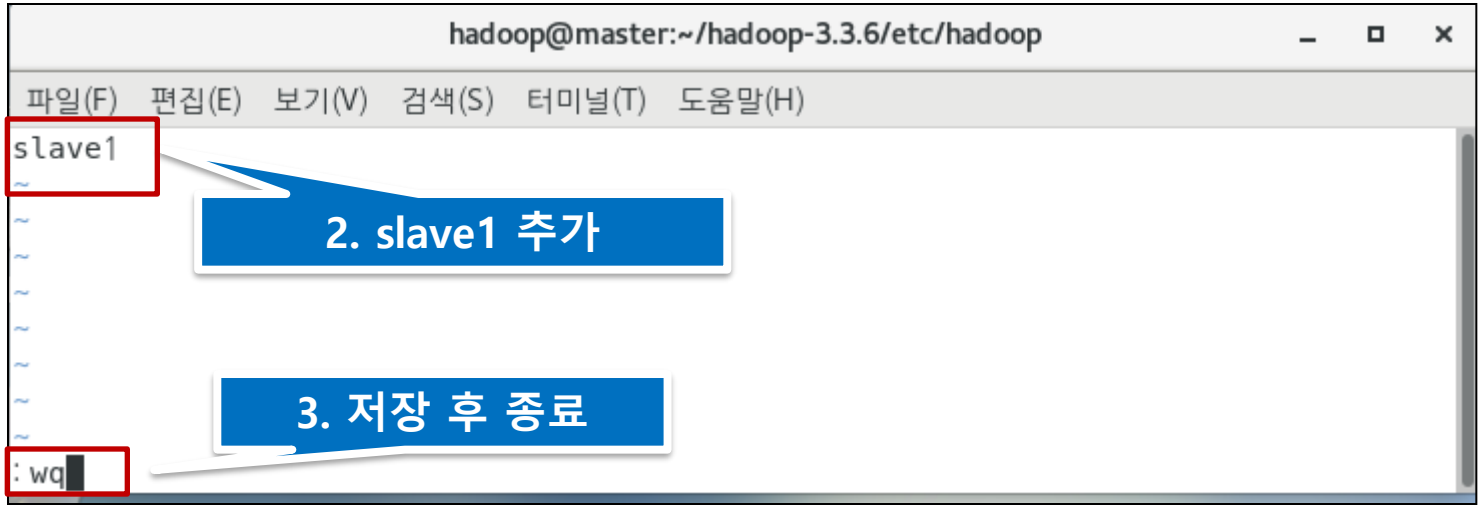
hadoop@master:~/hadoop-3.3.6/et…

7) Data node(Slave 서버) 지정

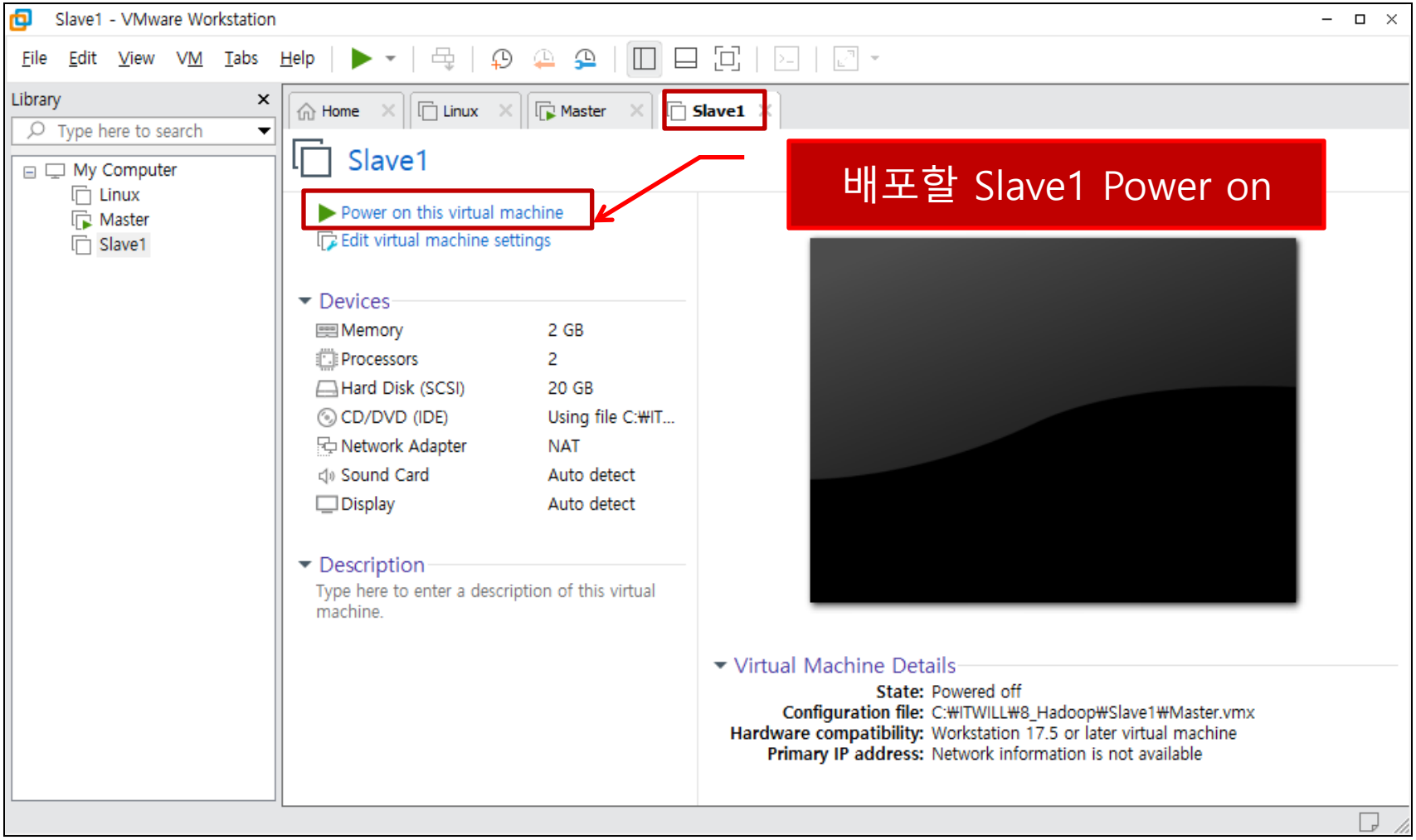
Multi-Node Cluster 모드용

[hadoop@master hadoop]$ **vi workers**

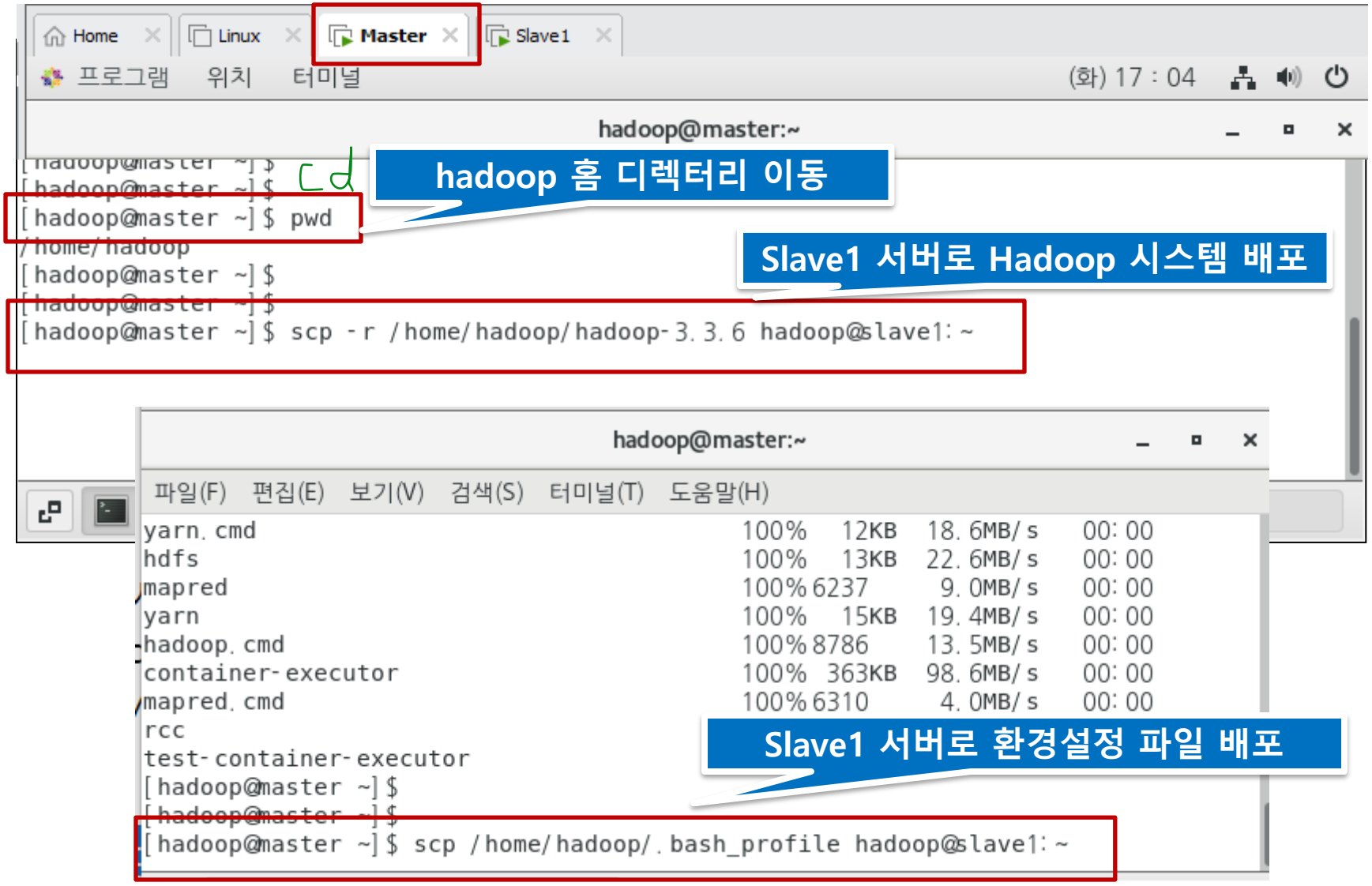
1. 설정파일 열기



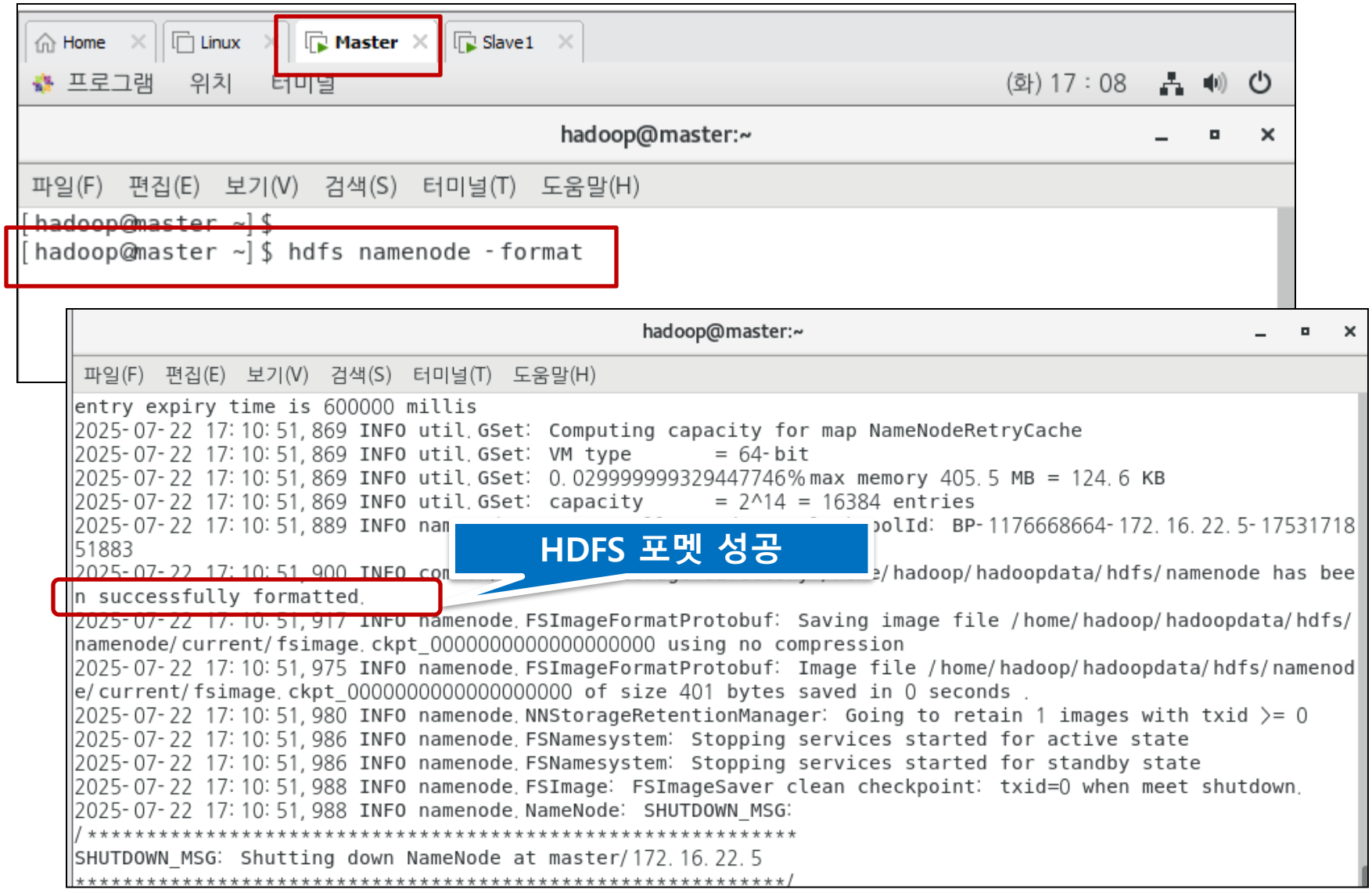
4. Hadoop 시스템 배포



4. Hadoop 시스템 배포(Master 작업)

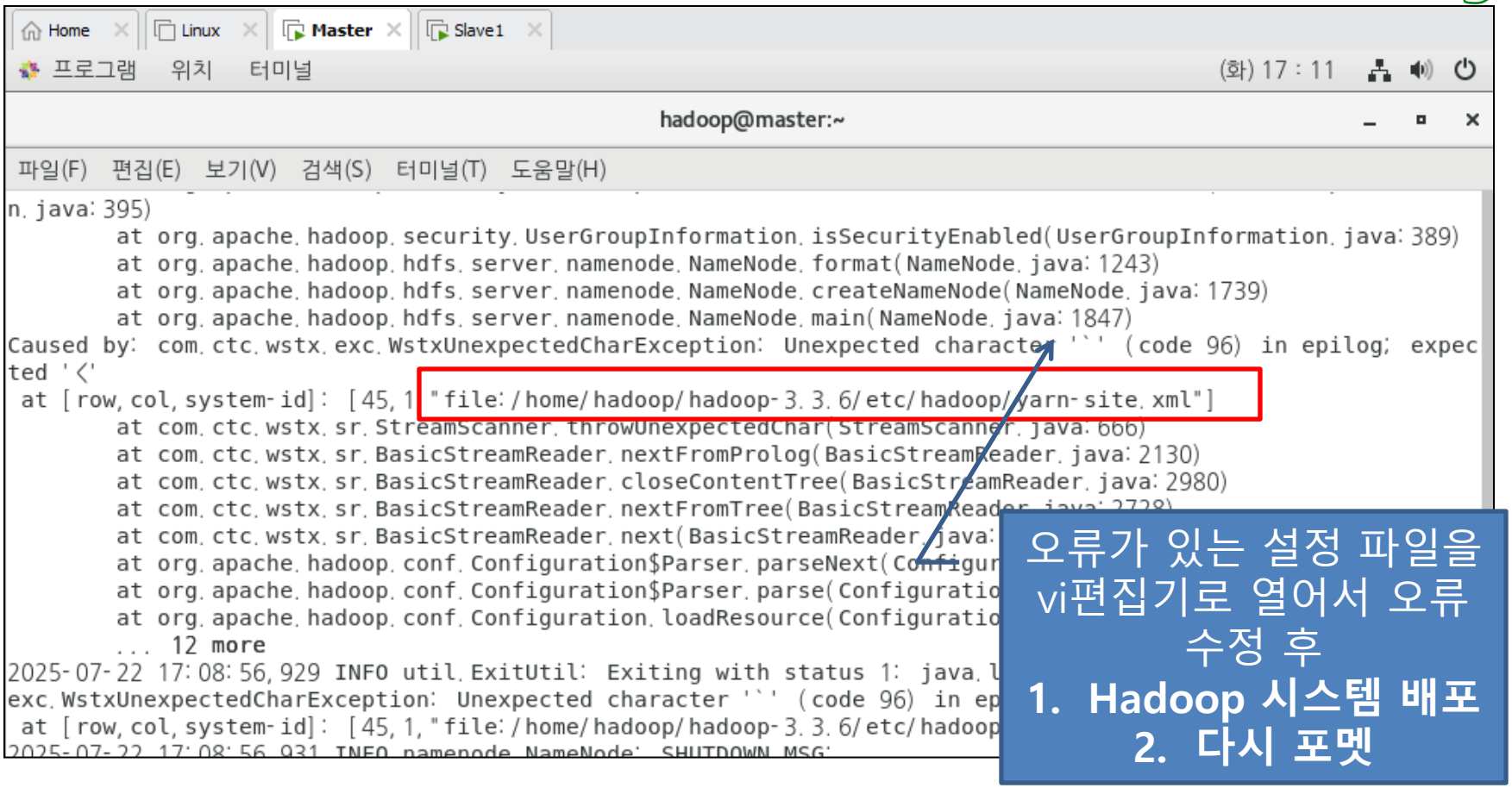


5. HDFS 포멧(최초 사용시)



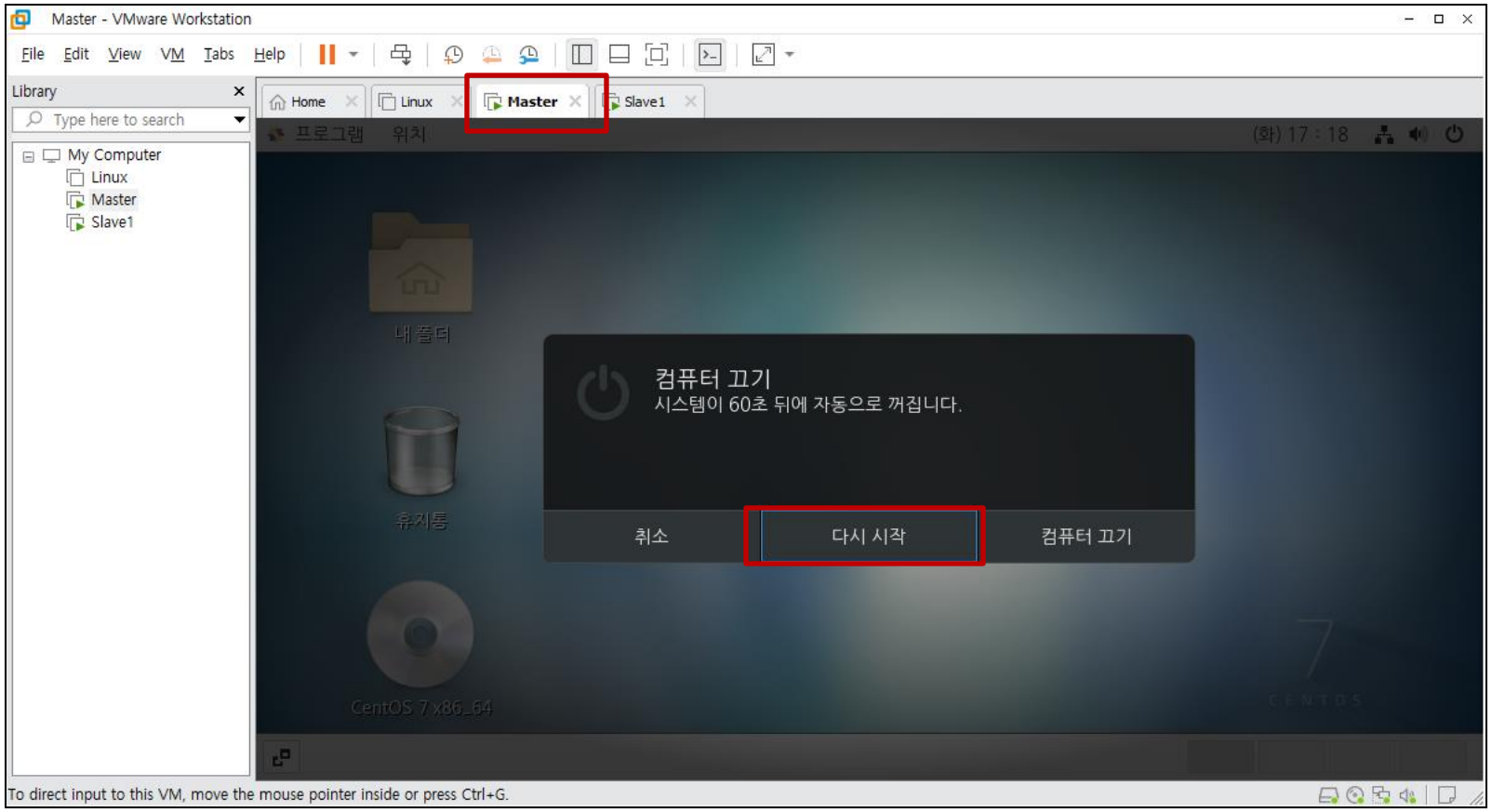
● 환경설정 파일에 오류가 있는 경우

오류나면 여기선행

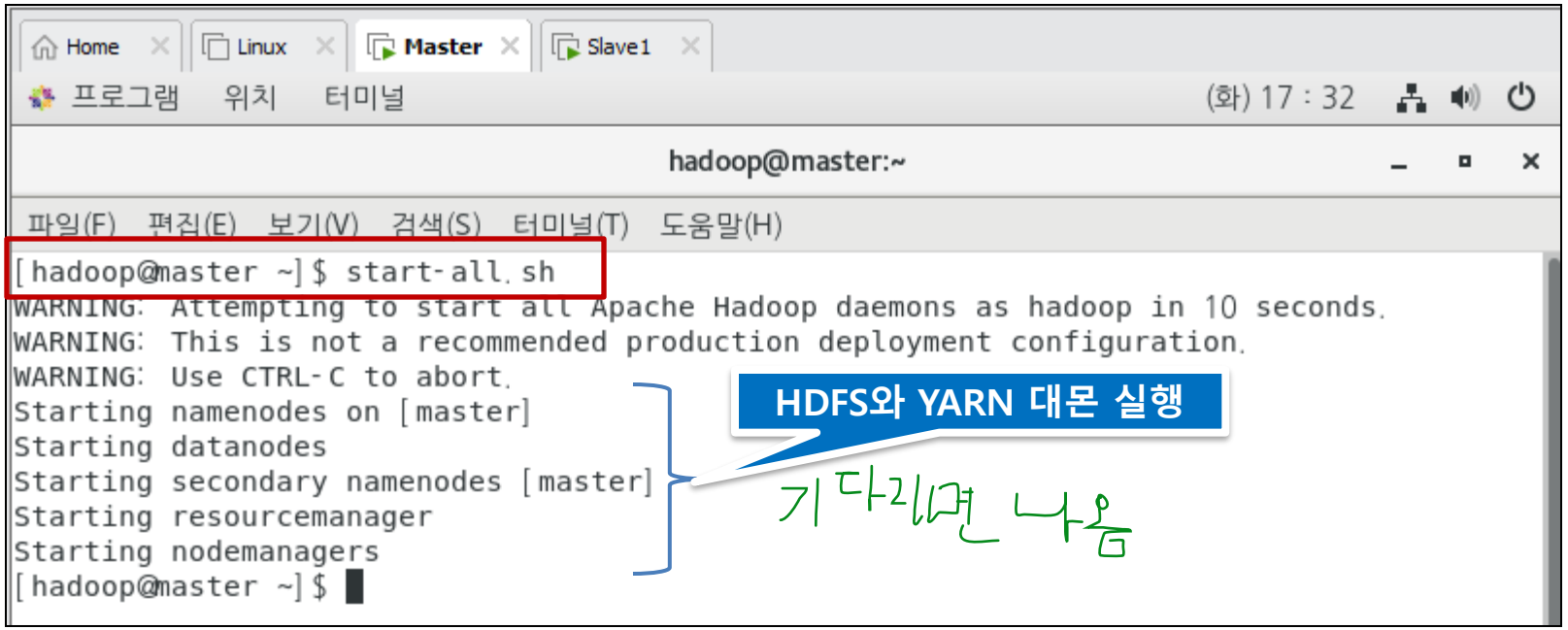


6. Hadoop/Yarn 데몬 시작

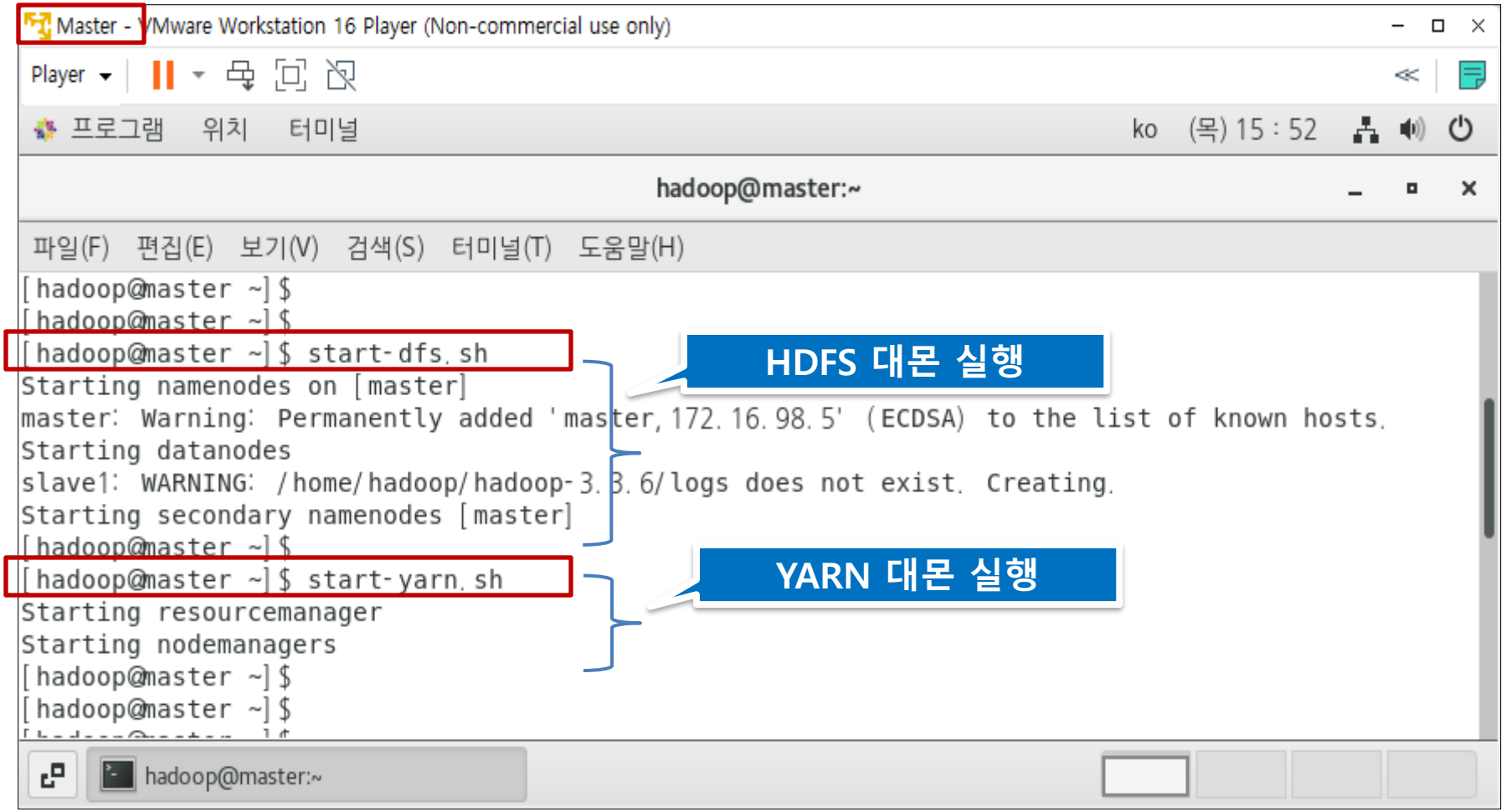
1) Master server 재시작



2) Hadoop/Yarn 동시 실행(Master에서 실행)

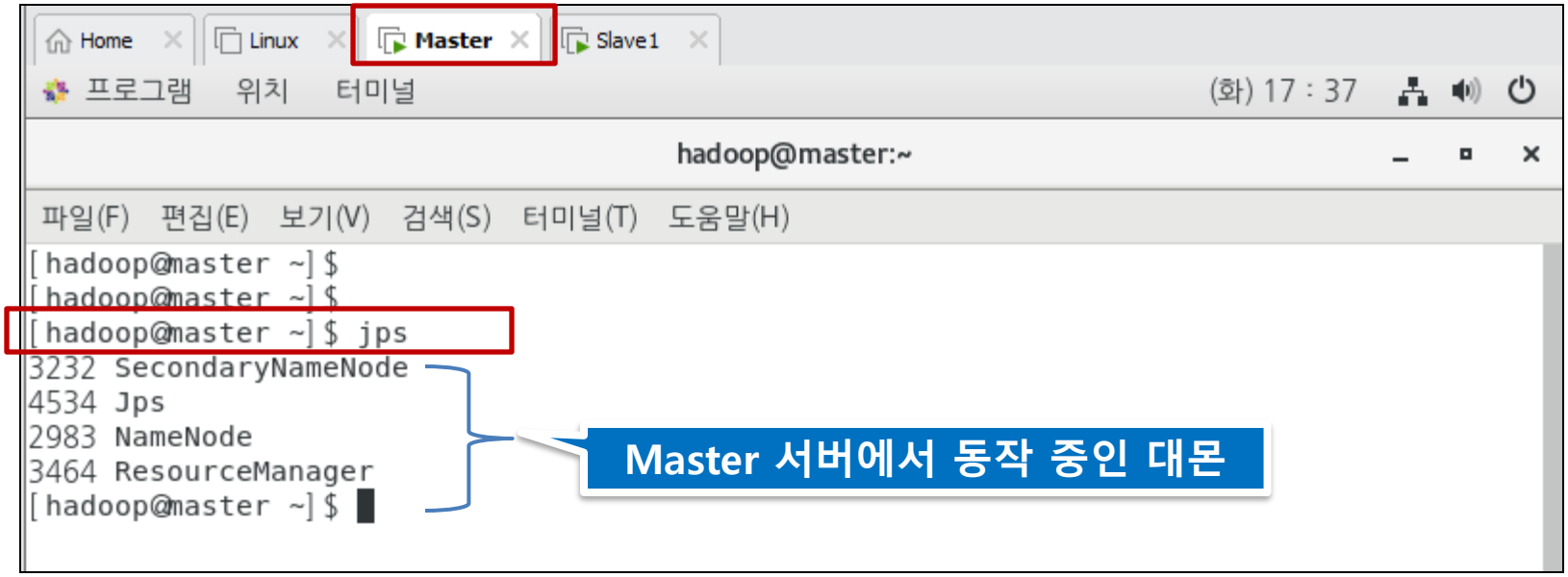


2) Hadoop과 Yarn 각각 실행(Master에서 실행)



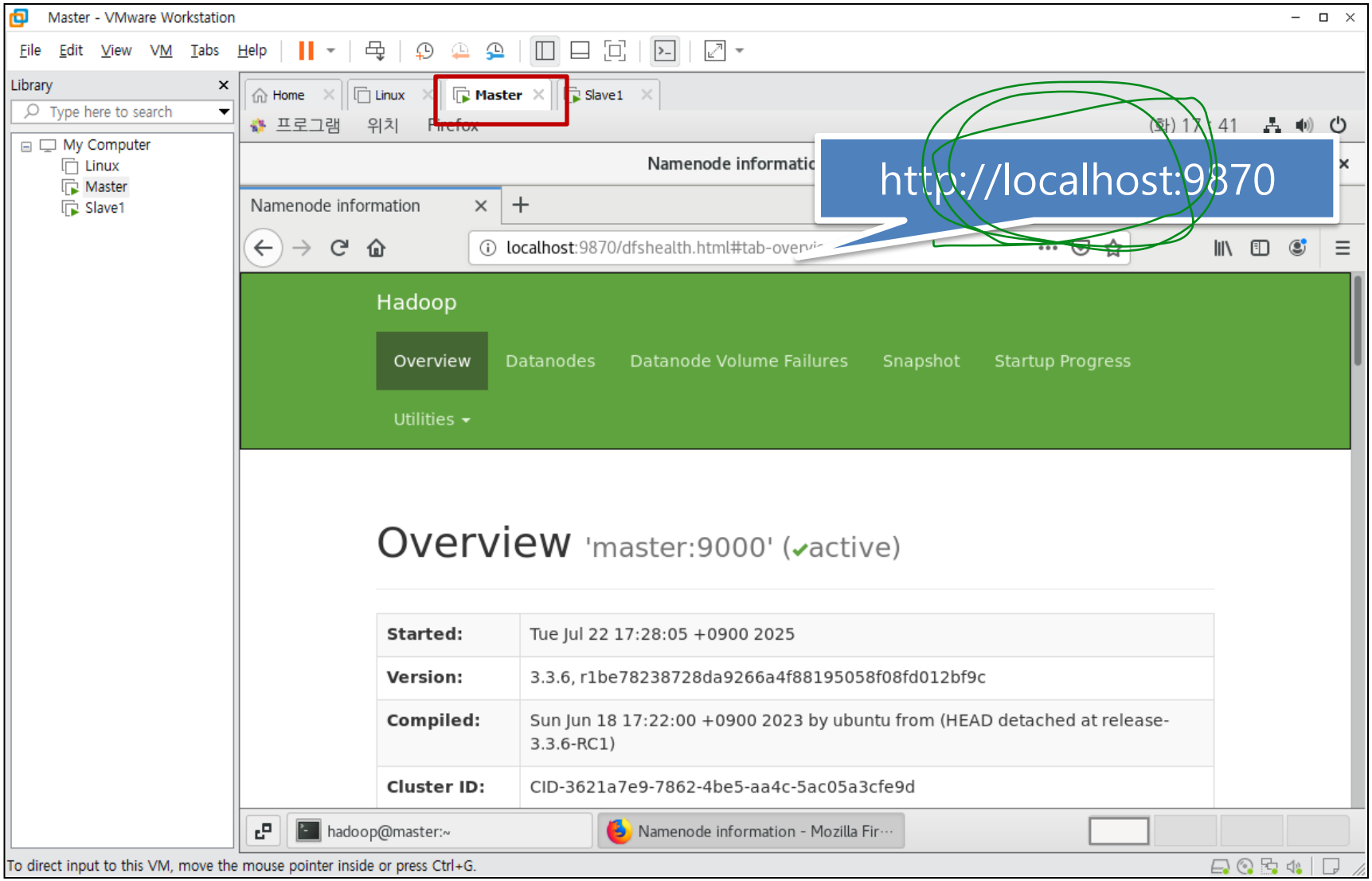
7. Hadoop system 상태 확인

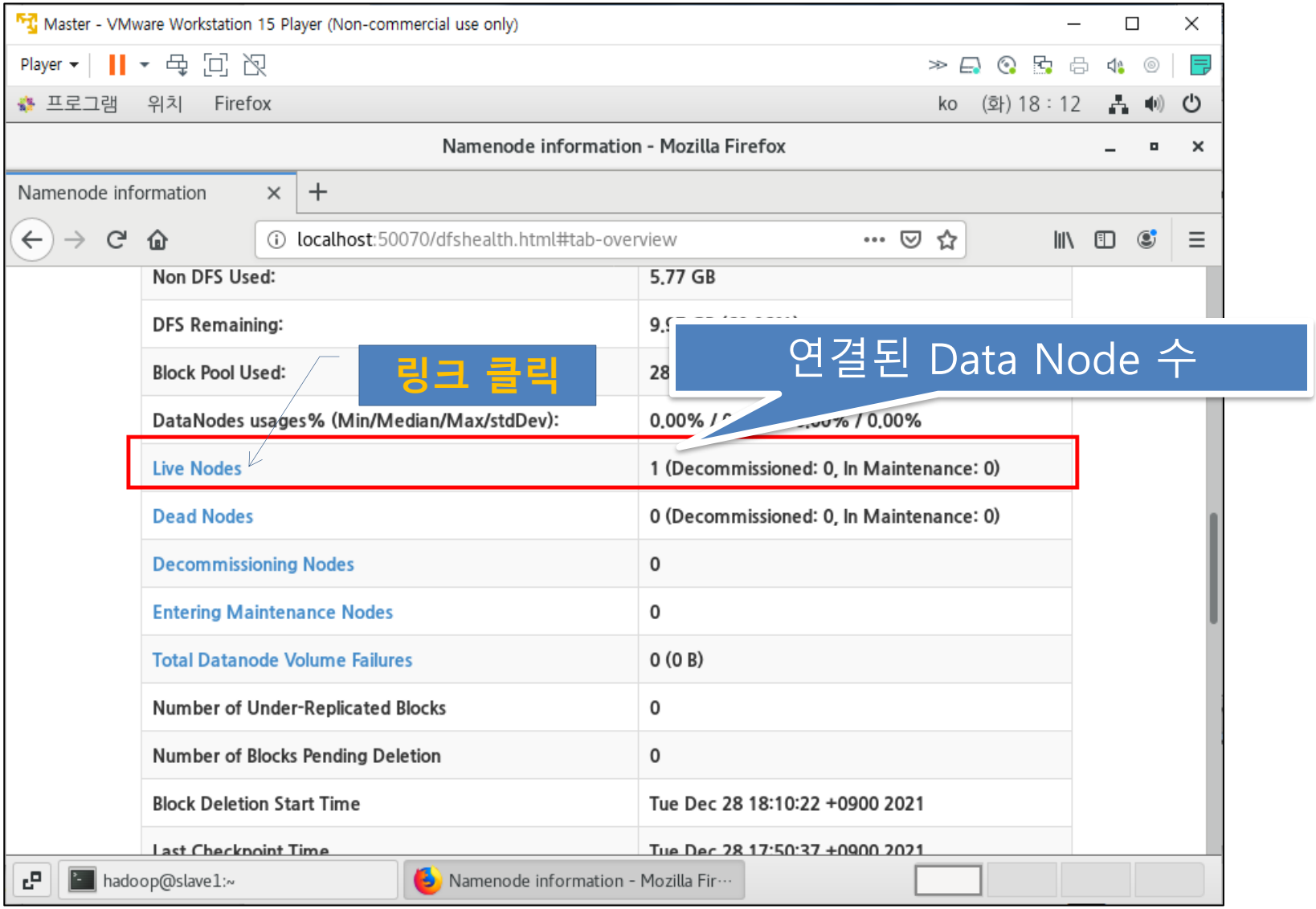
❖ Master(NameNode서버) : NameNode, ResourceManager 데몬 실행



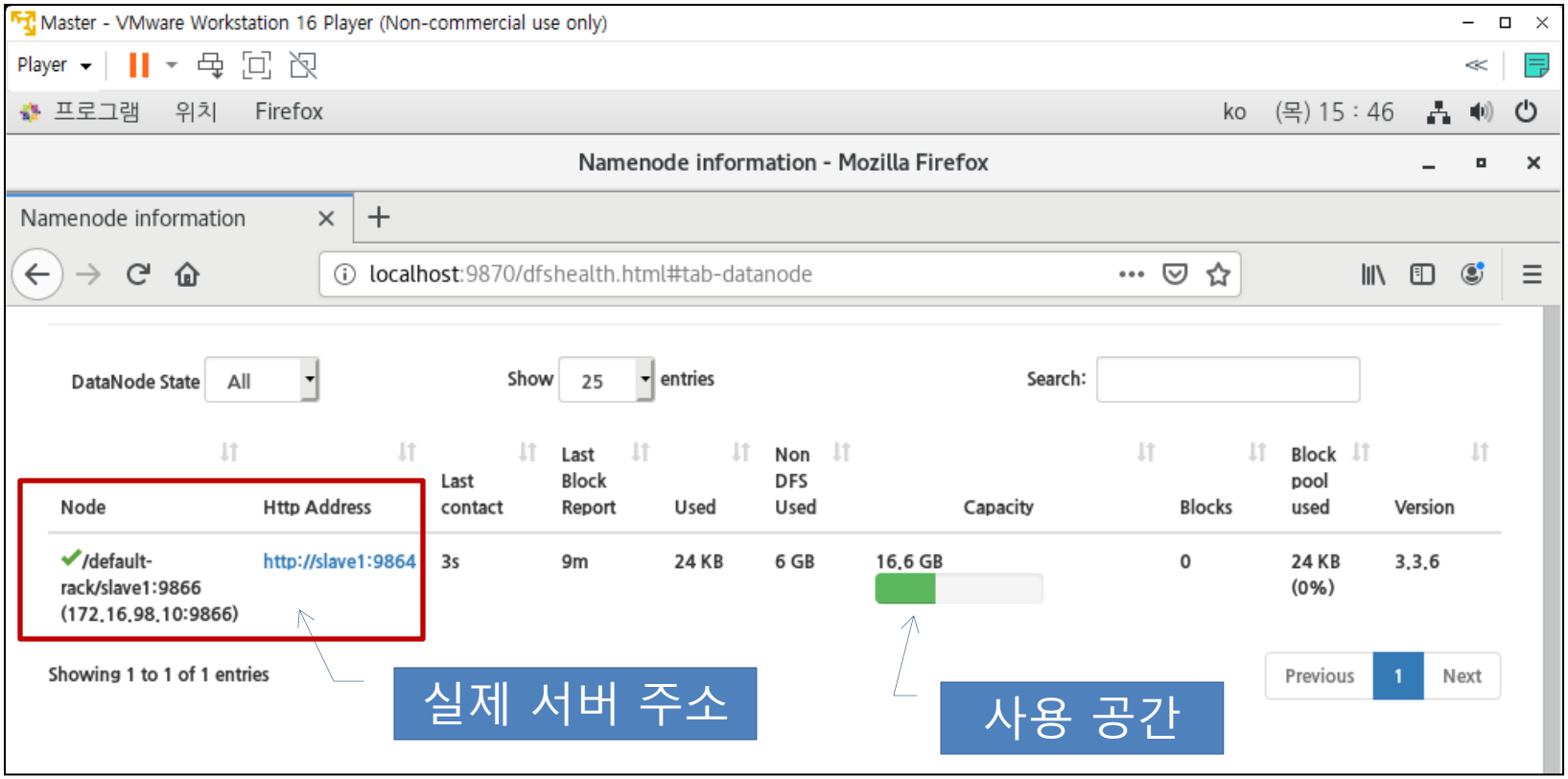
❖ Slave(DataNode서버): DataNode, NodeManager데몬 실행

Web에서 Hadoop 상태 확인

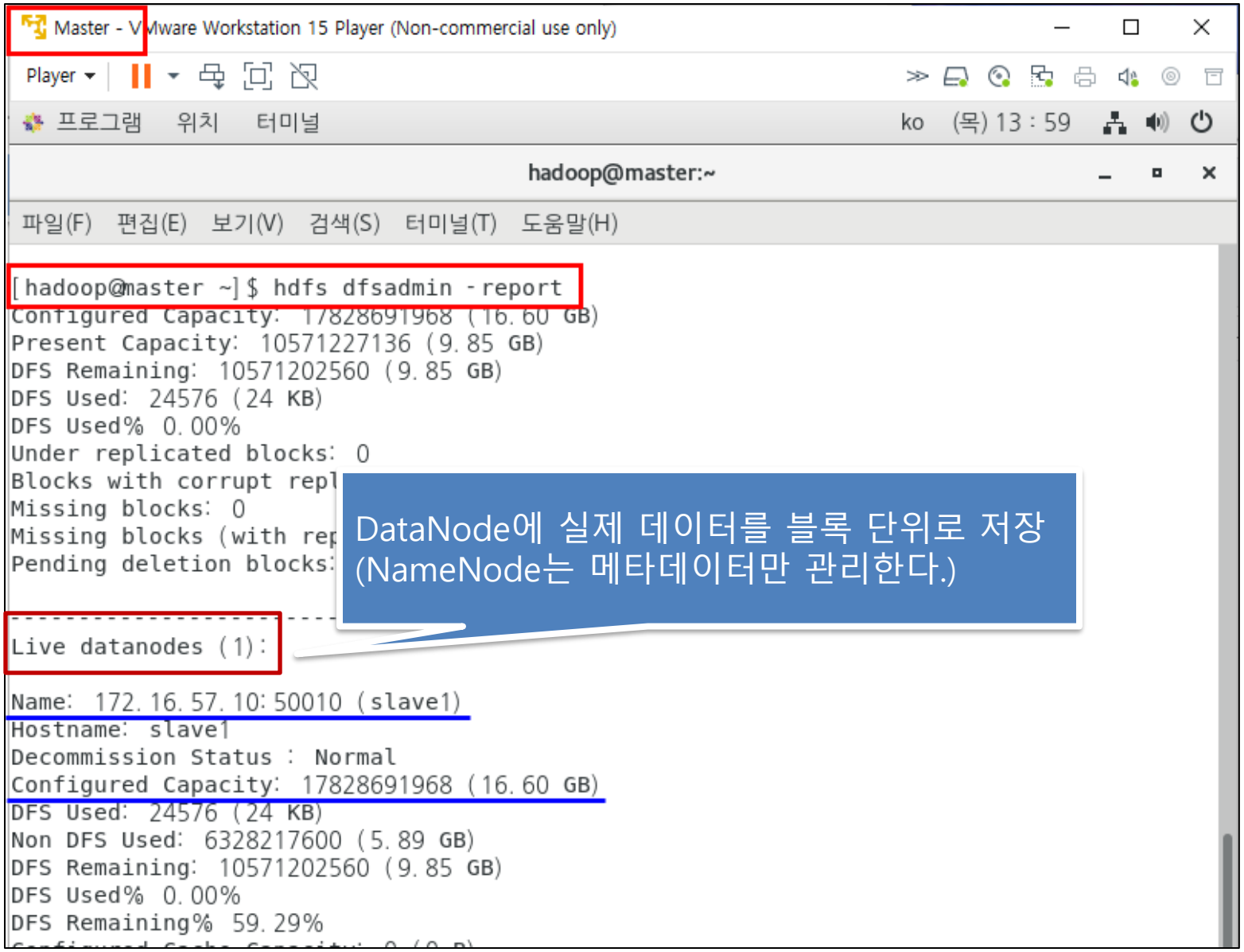



Master - VMware Workstation 15 Player (Non-commercial use only)
Player
프로그램 위치 Firefox
ko (화) 18 : 12
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-overview
Non DFS Used: 5.77 GB
DFS Remaining:
Block Pool Used:
DataNodes usages% (Min/Median/Max/stdDev): 0.00% / 0.00%
Live Nodes 1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes 0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes 0
Entering Maintenance Nodes 0
Total Datanode Volume Failures 0 (0 B)
Number of Under-Replicated Blocks 0
Number of Blocks Pending Deletion 0
Block Deletion Start Time Tue Dec 28 18:10:22 +0900 2021
링크 클릭
연결된 Data Node 수
hadoop@slave1:~
Namenode information - Mozilla Fir…

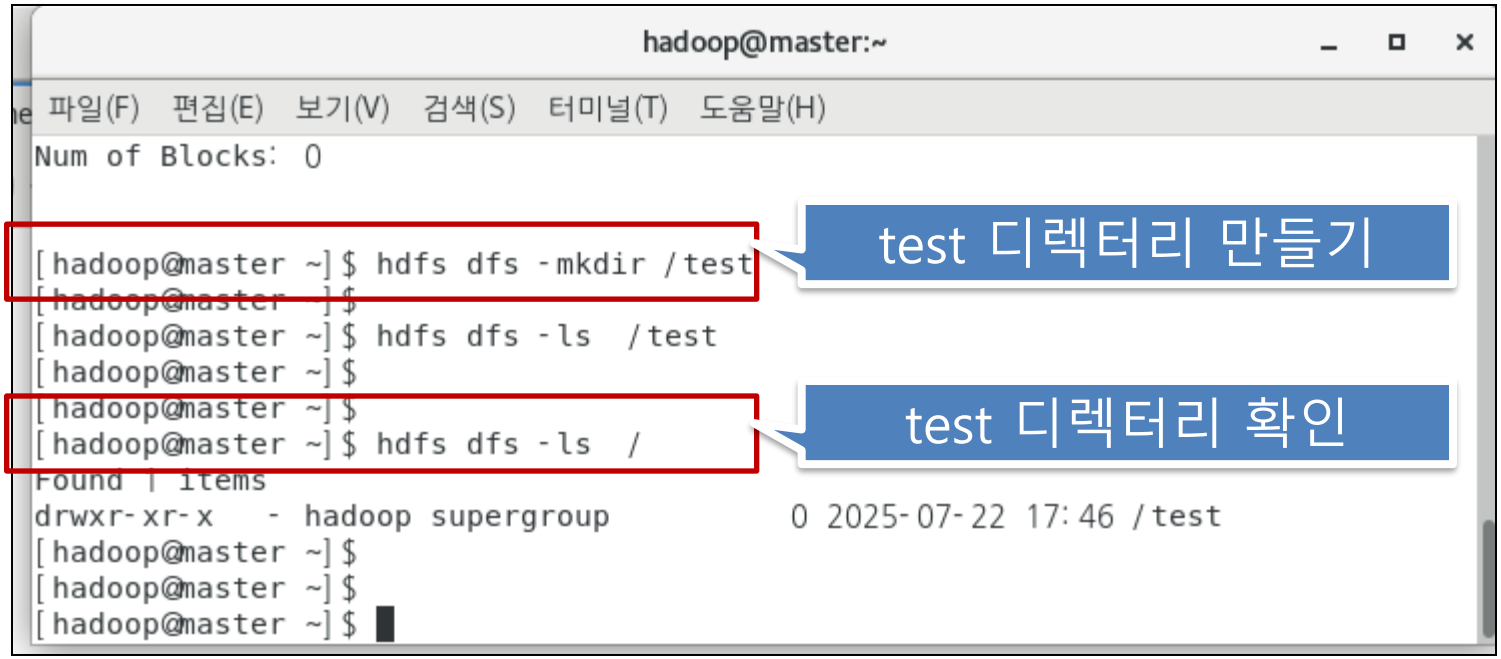
- 선택된 Data Node 정보 및 사용 내용



HDFS(하둡 저장소) 보기



[실습] HDFS에 디렉터리 생성 & 확인



8. Hadoop/Yarn 데몬 종료