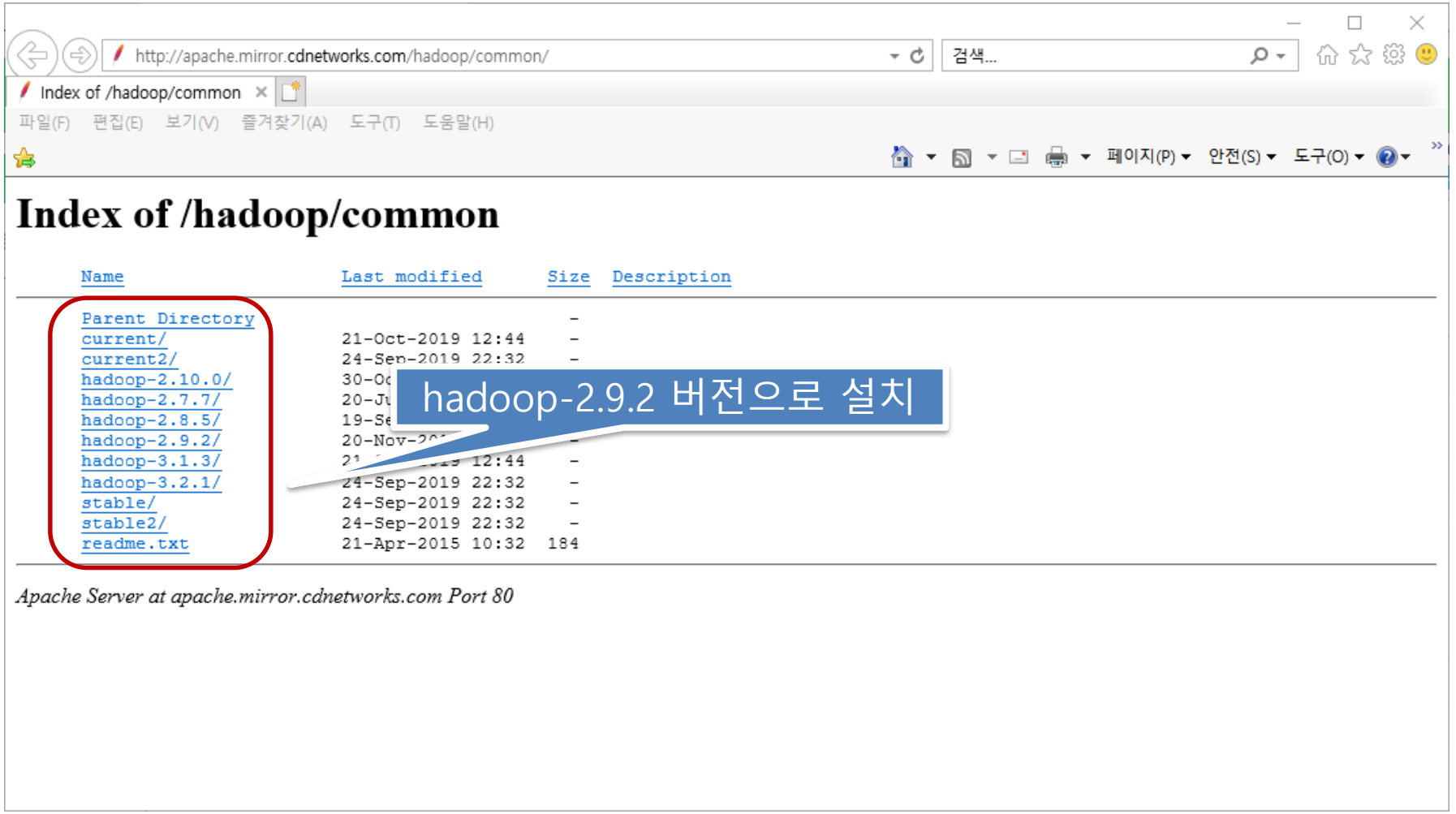
# 3. Hadoop 설치

## 목 차

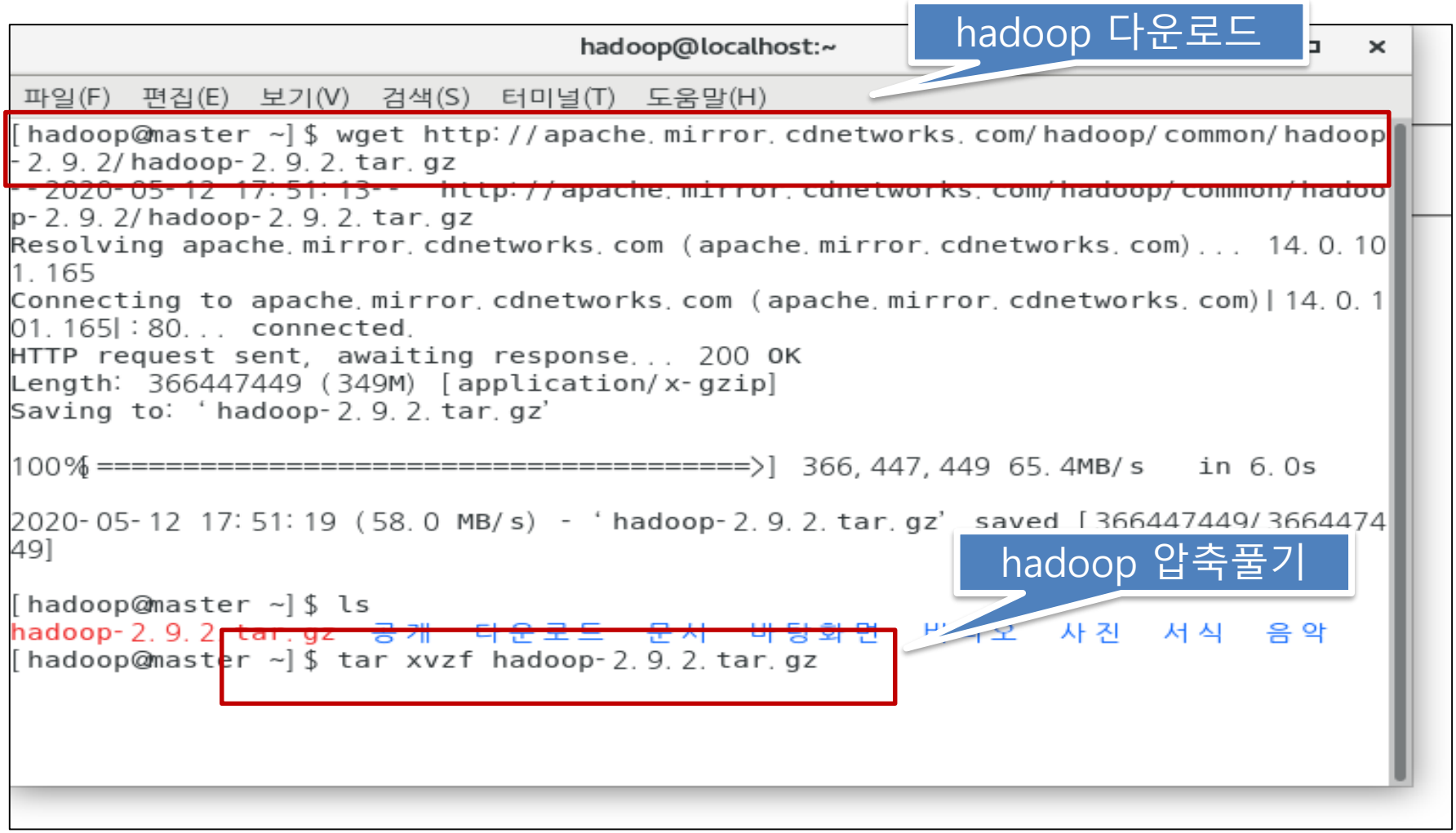
# 1. Hadoop 설치(다운로드 가능 버전 확인)

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

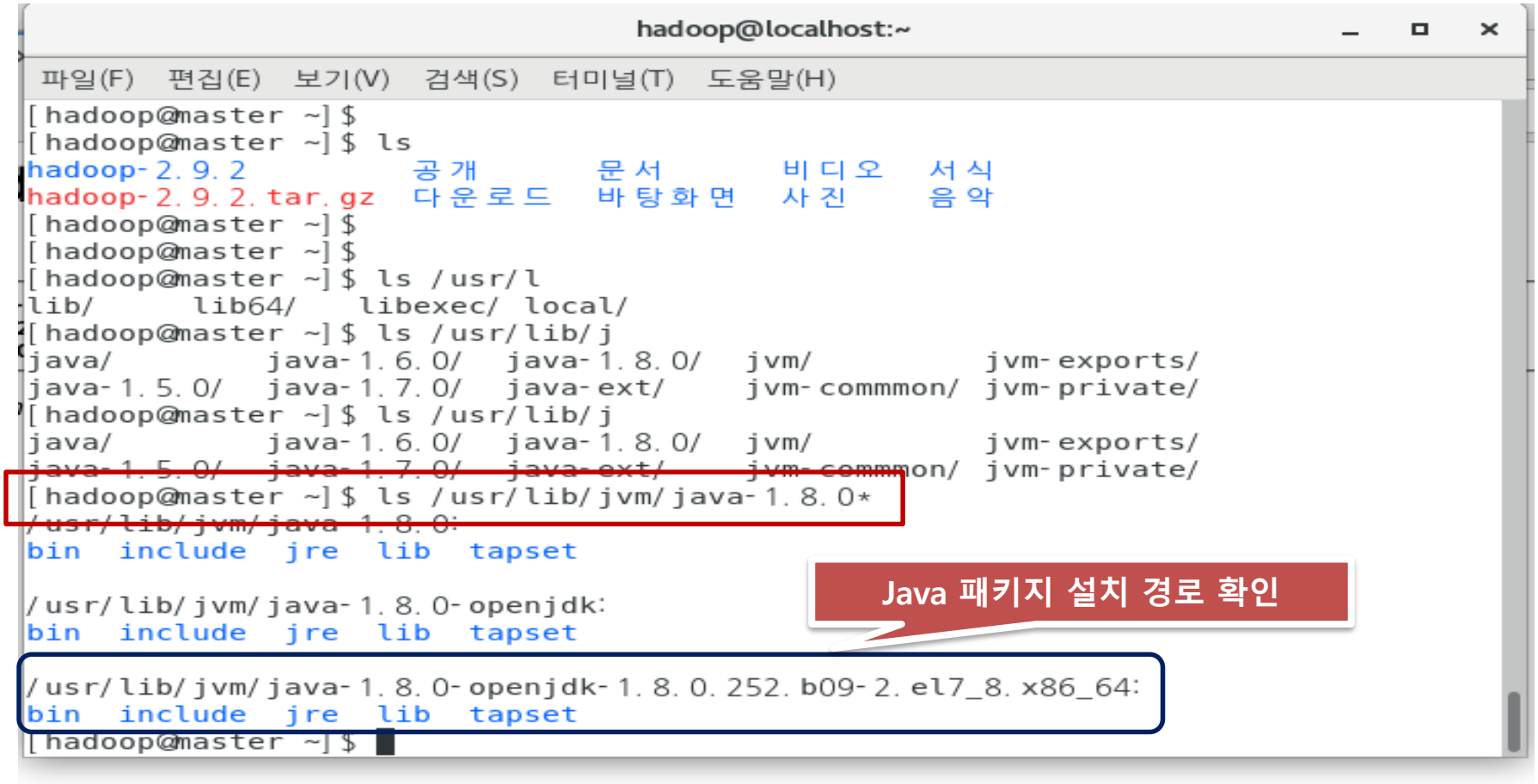


# 1. Hadoop 설치(다운로드 가능 버전 확인)

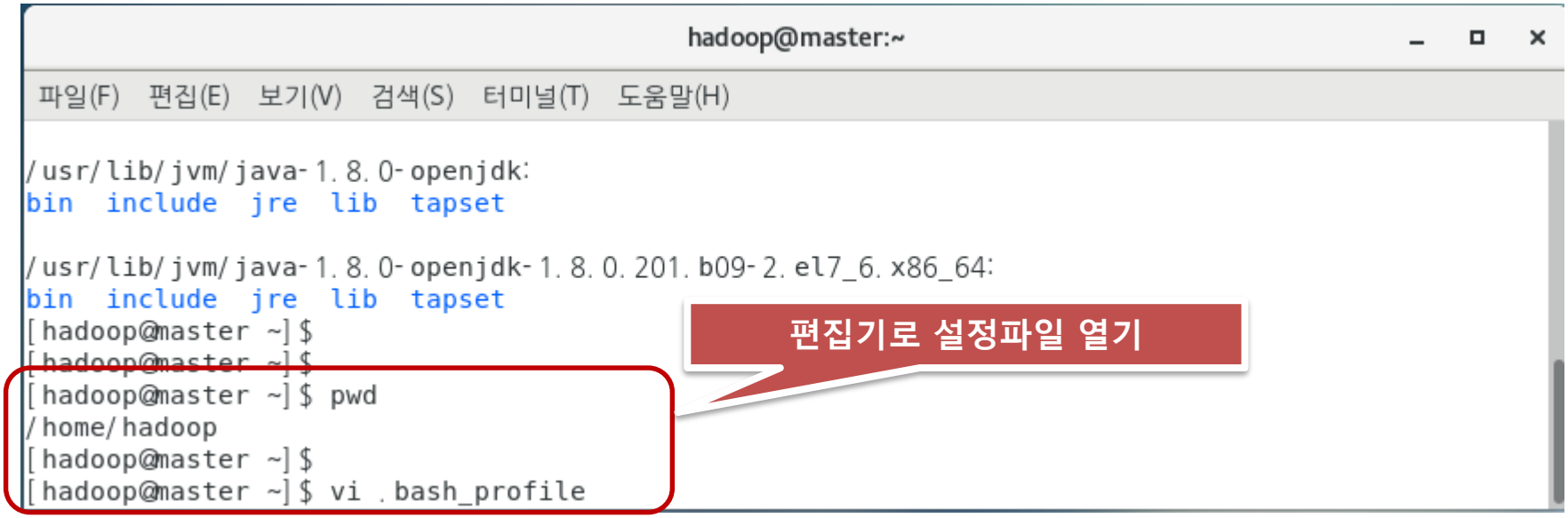
http://apache.mirror.cdnetworks.com/hadoop/common/ 에서 다운로드 가능한 버전 확인



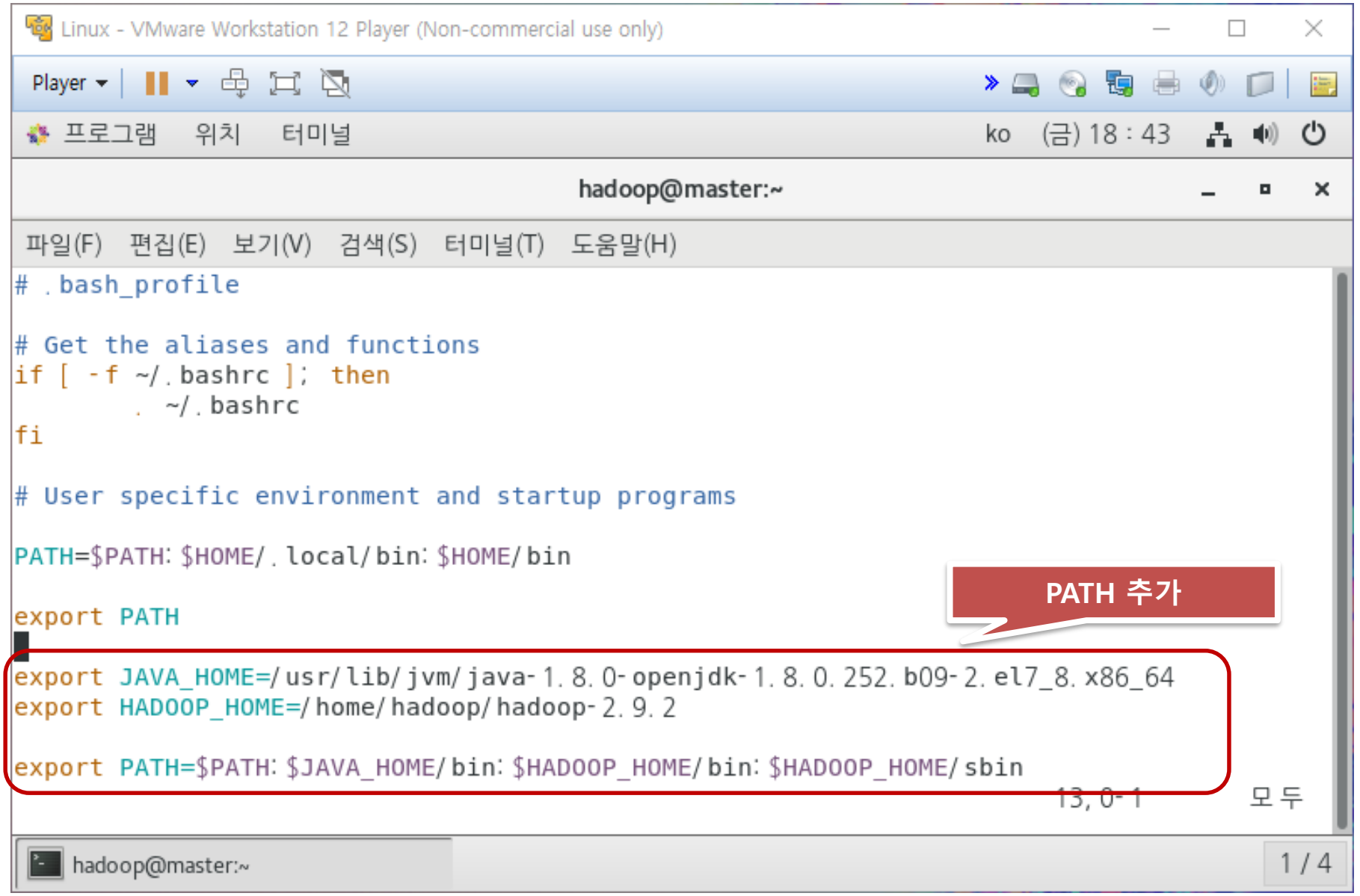
# 2. Java 홈/ Hadoop 홈 Path 설정



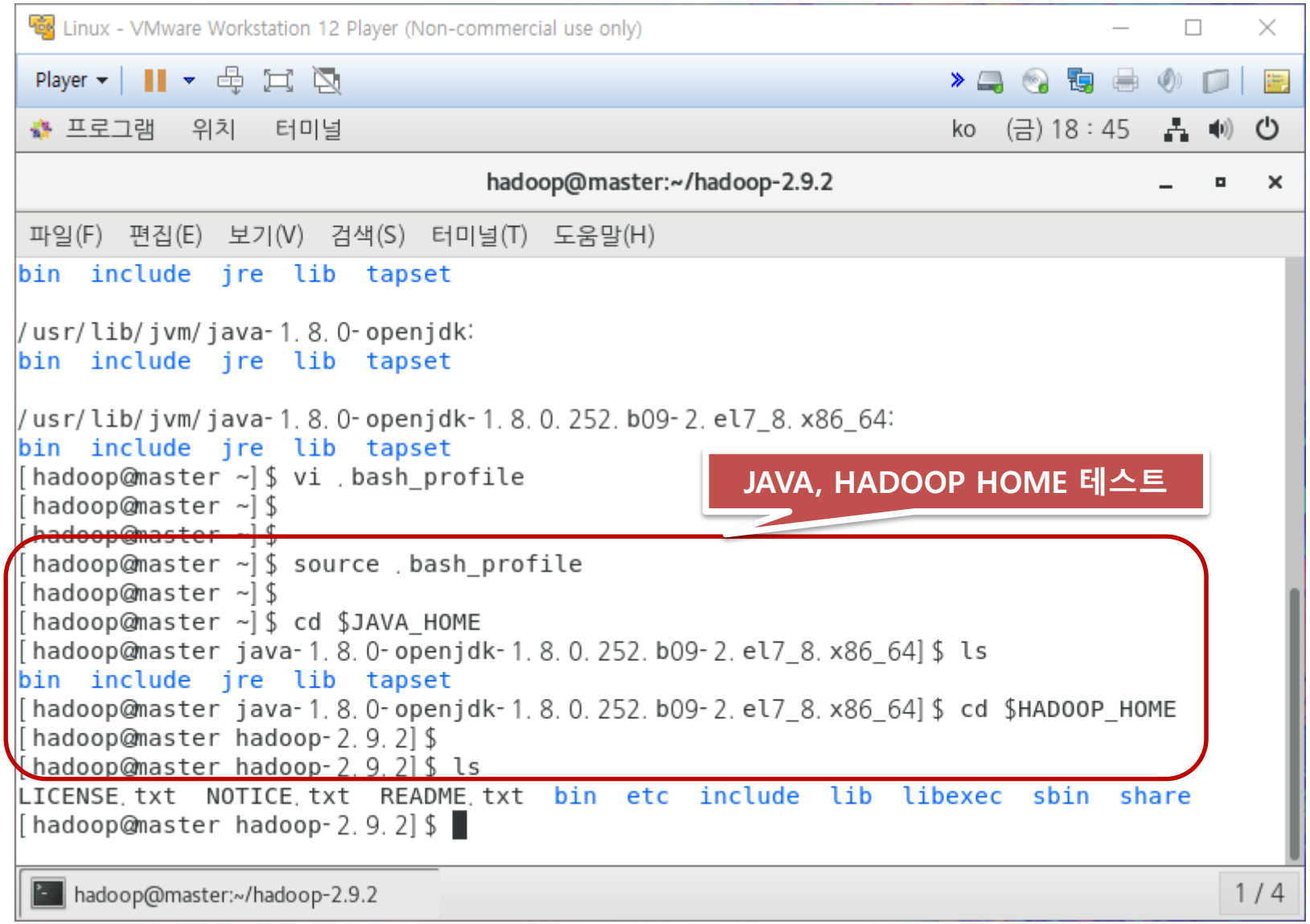
# .bash_profile 설정 : Hadoop & Java 홈 디렉터리 지정



# .bash_profile 설정 : Hadoop & Java 홈 디렉터리 지정



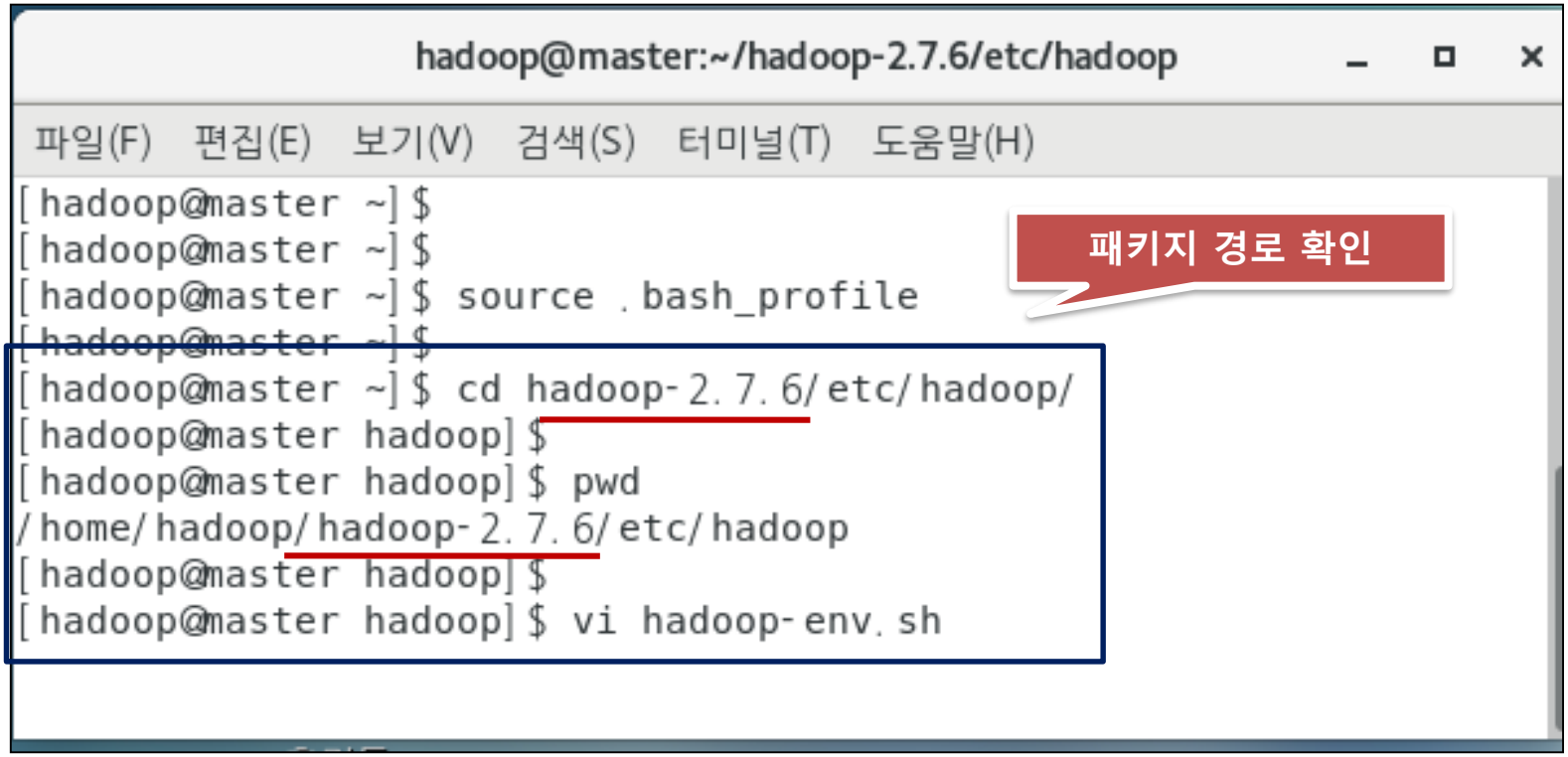
# .bash_profile 적용/환경변수 확인



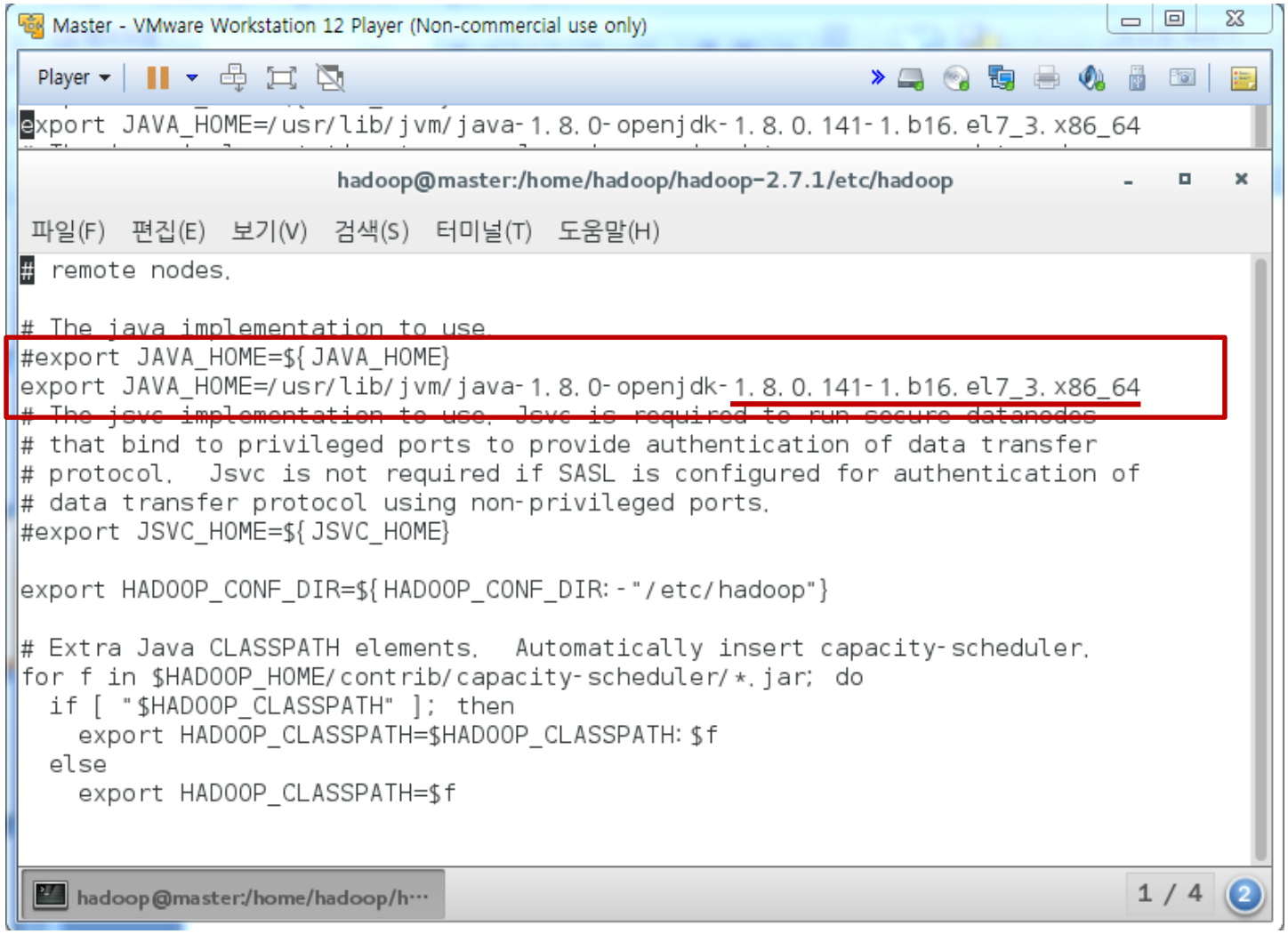
# 3. Hadoop 환경 설정 관련 파일

| 파일명 | 설정 내용 |
|---|---|
| **hadoop-env.sh** | Hadoop을 실행하는 설정파일[JDK경로, Class 패스, 데몬 실행 옵션 등] |
| **yarn-env.sh** | YARN(Mapreduce ver2)을 실행하는 설정파일[JDK경로 등] |
| masters | 보조 Name node를 실행 할 서버 설정 |
| **slaves** | Data node를 실행할 서버 설정 |
| **core-site.xml** | 분산파일시스템[HDFS]를 지정하고, HDFS와 YARN에서 이용될 임시 디렉터리 지정 |
| **hdfs-site.xml** | Name node와 Data node의 디렉터리 지정 |
| **mapred-site.xml** | Mapreduce에서 사용할 환경설정 파일 |
| **yarn-site.xml** | YARN[Hadoop 클러스터 관리 시스템] 설정파일, Resource Manager, Node Manager 지정 |

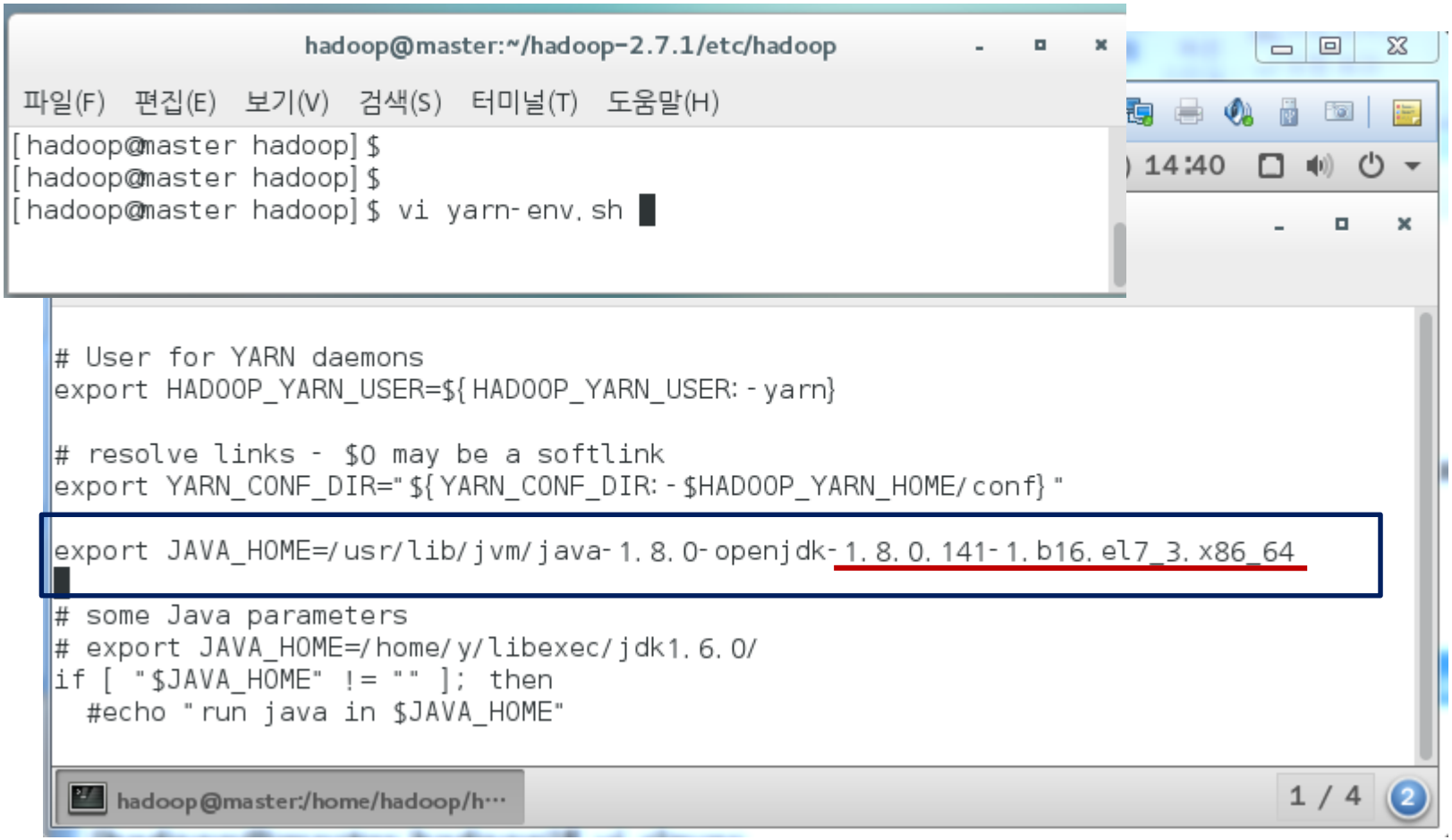
# Hadoop 환경 설정


hadoop@master:~/hadoop-2.7.6/etc/hadoop
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master ~]$
[hadoop@master ~]$
[hadoop@master ~]$ source .bash_profile
[hadoop@master ~]$
[hadoop@master ~]$ cd hadoop-2.7.6/etc/hadoop/
[hadoop@master hadoop]$
[hadoop@master hadoop]$ pwd
/home/hadoop/hadoop-2.7.6/etc/hadoop
[hadoop@master hadoop]$
[hadoop@master hadoop]$ vi hadoop-env.sh
패키지 경로 확인

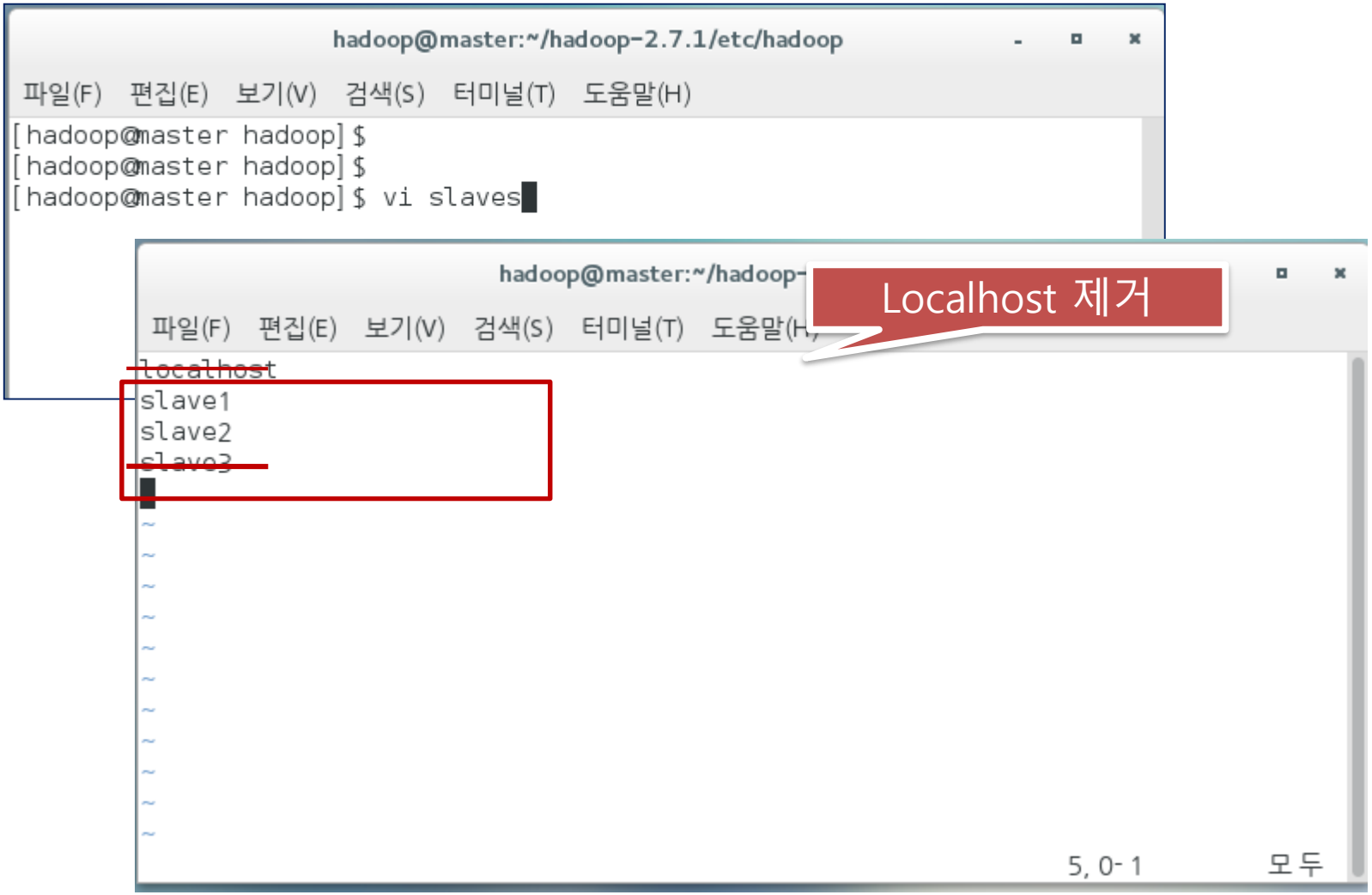
# Hadoop-env.sh 설정(jdk 추가)



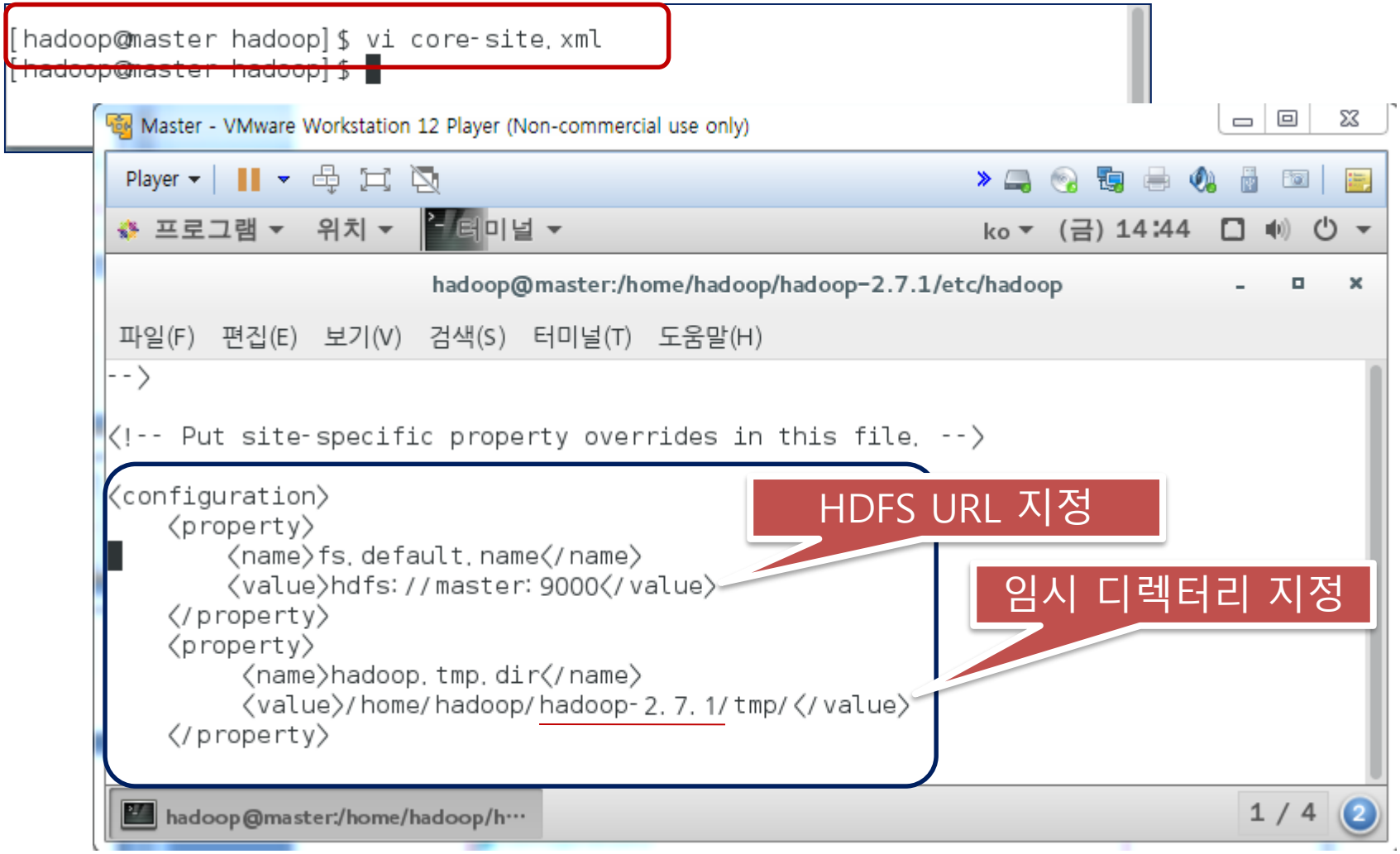
# Yarn 환경 설정(jdk 추가)



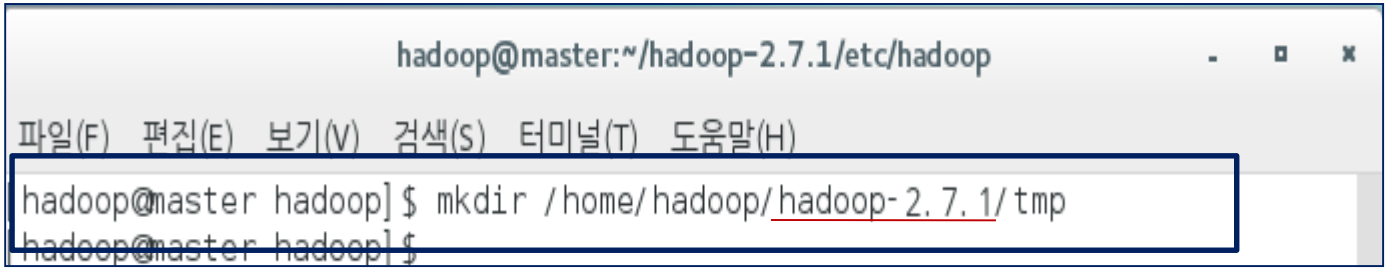
# Data node(Slave 서버) 지정



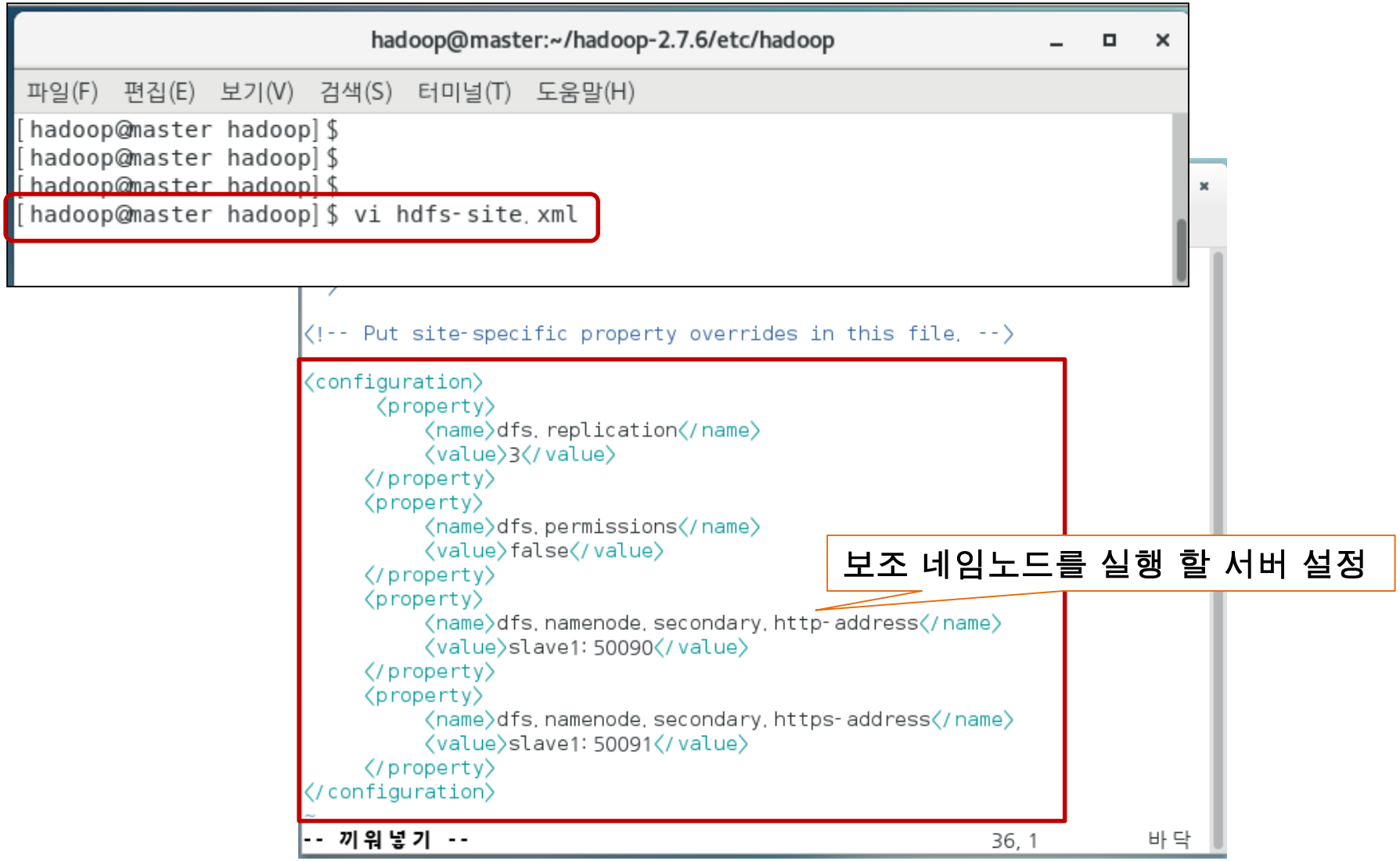
# HDFS의 Name node 지정, HDFS 임시 디렉터리 지정



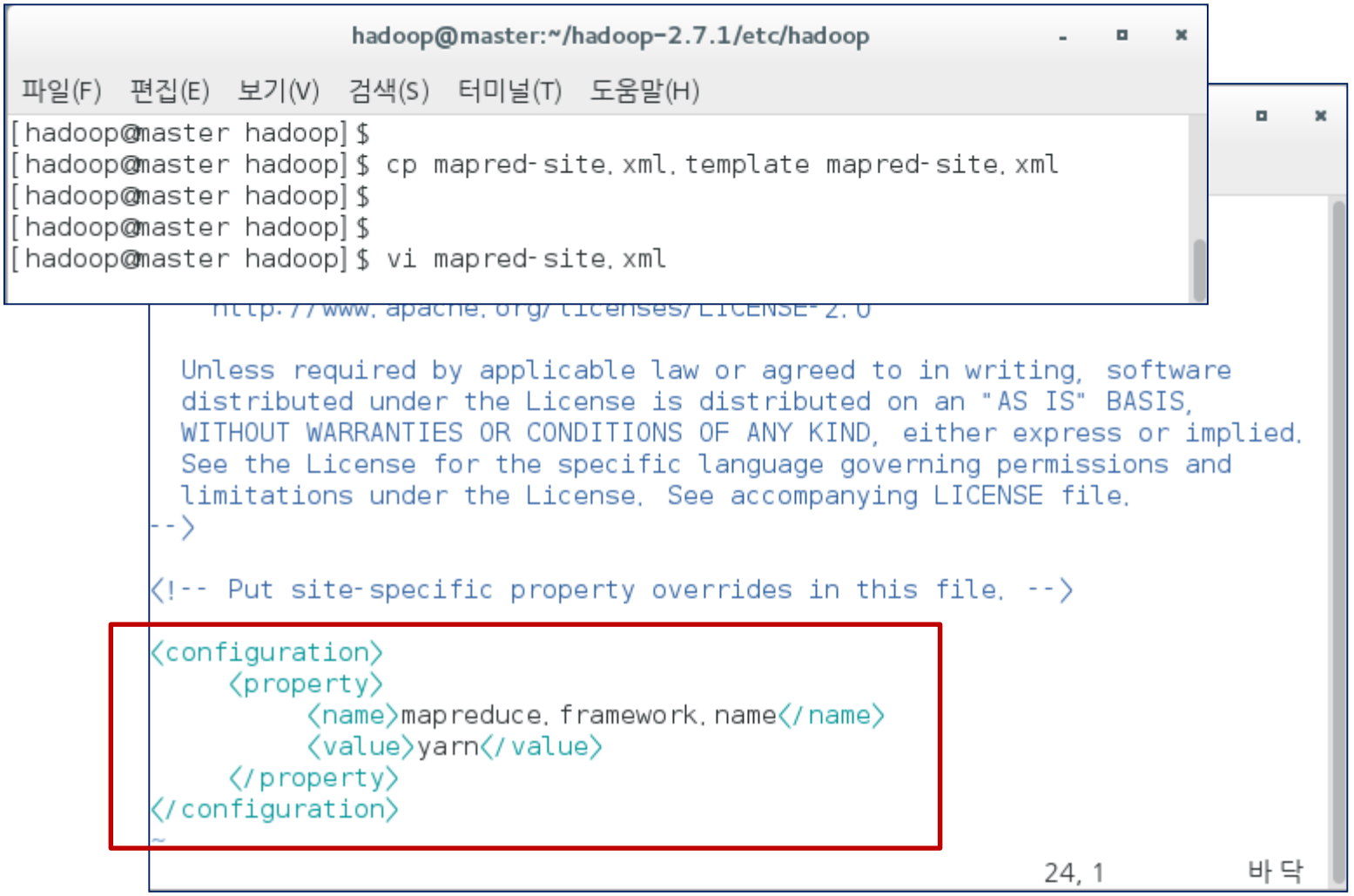
# HDFS에 tmp 임시 데이터용 디렉터리 생성 (core-site.xml 에서 지정한 디렉터리)



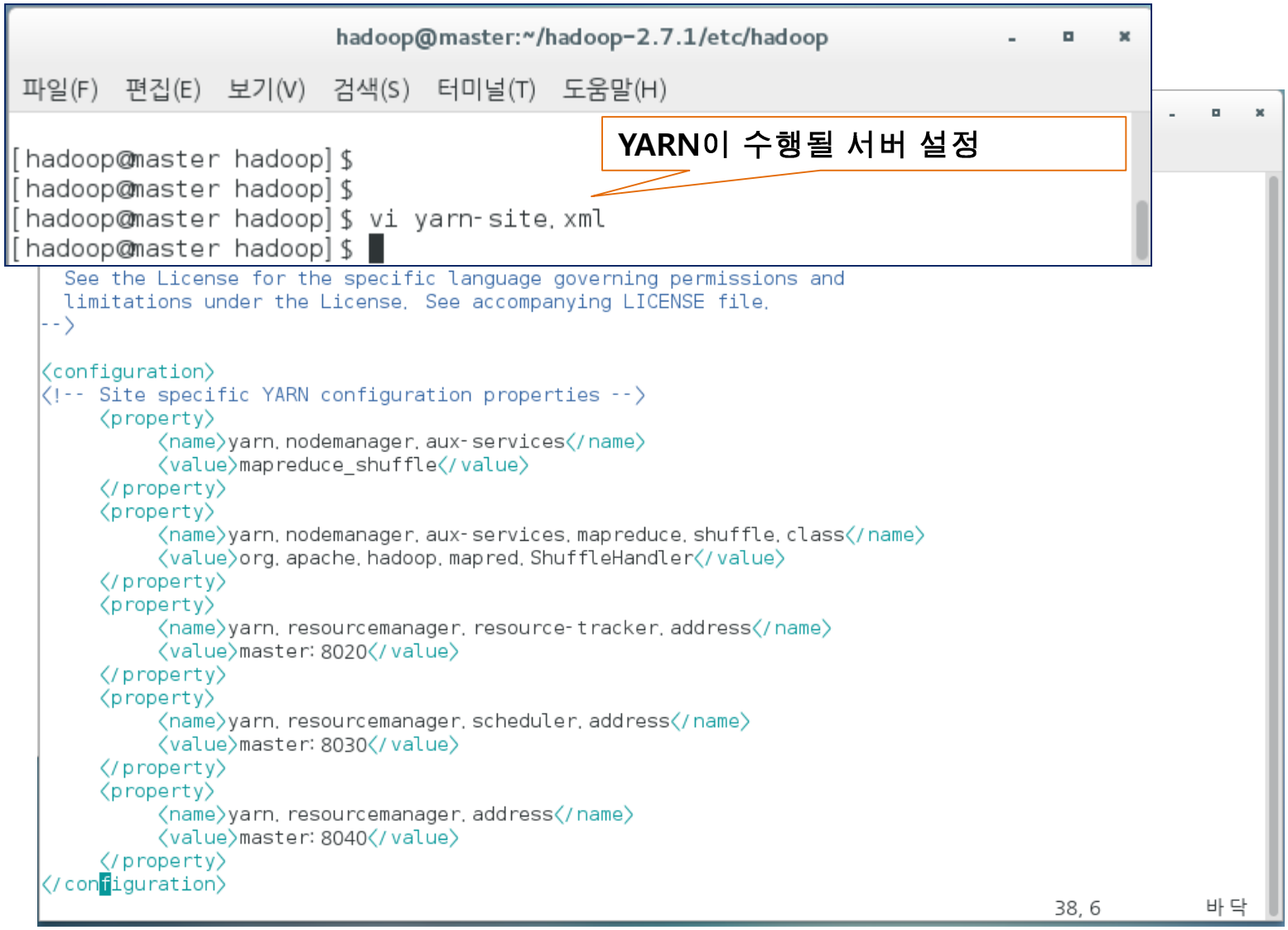
# HDFS 관련 서버 설정


hadoop@master:~/hadoop-2.7.6/etc/hadoop
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master hadoop]$
[hadoop@master hadoop]$
[hadoop@master hadoop]$
[hadoop@master hadoop]$ vi hdfs-site.xml
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>dfs.replication</name>
<value>3</value>
</property>
<property>
<name>dfs.permissions</name>
<value>false</value>
</property>
<property>
<name>dfs.namenode.secondary.http-address</name>
<value>slave1:50090</value>
</property>
<property>
<name>dfs.namenode.secondary.https-address</name>
<value>slave1:50091</value>
</property>
</configuration>
-- 끼워넣기 --
36,1
바닥
보조 네임노드를 실행 할 서버 설정

# MapReduce 애플리케이션 정보 설정



# Hadoop 관리 시스템 설정
# (Resource Manager, Node Manager)



# 4. Hadoop 시스템 배포

**1) Hadoop** 디렉터리 배포

[hadoop@master hadoop]$ cd

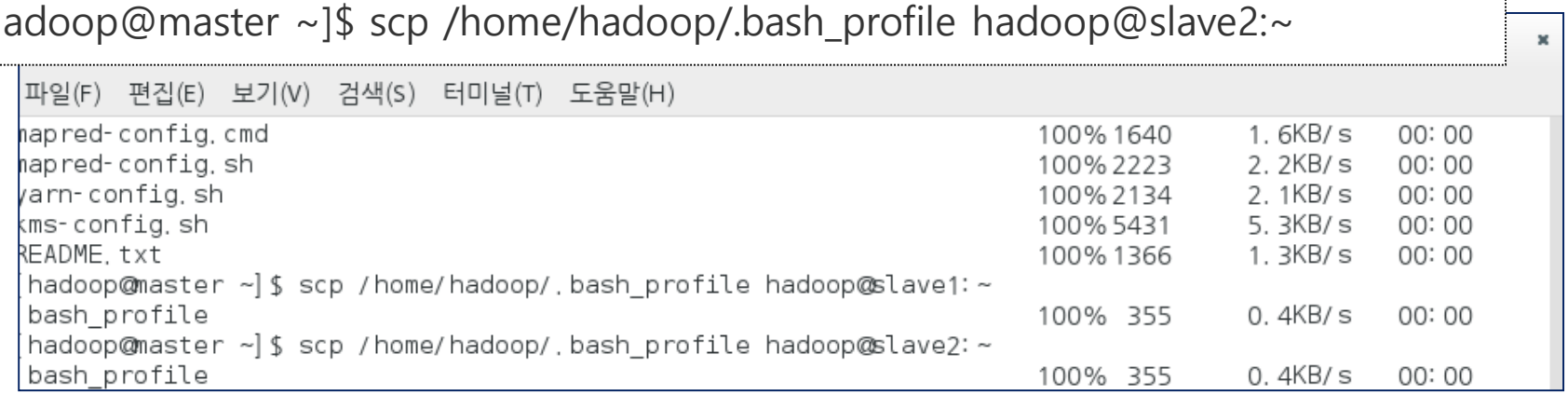
[hadoop@master ~]$ scp -r /home/hadoop/hadoop-2.7.1 hadoop@slave1:~

[hadoop@master ~]$ scp -r /home/hadoop/hadoop-2.7.1 hadoop@slave2:~

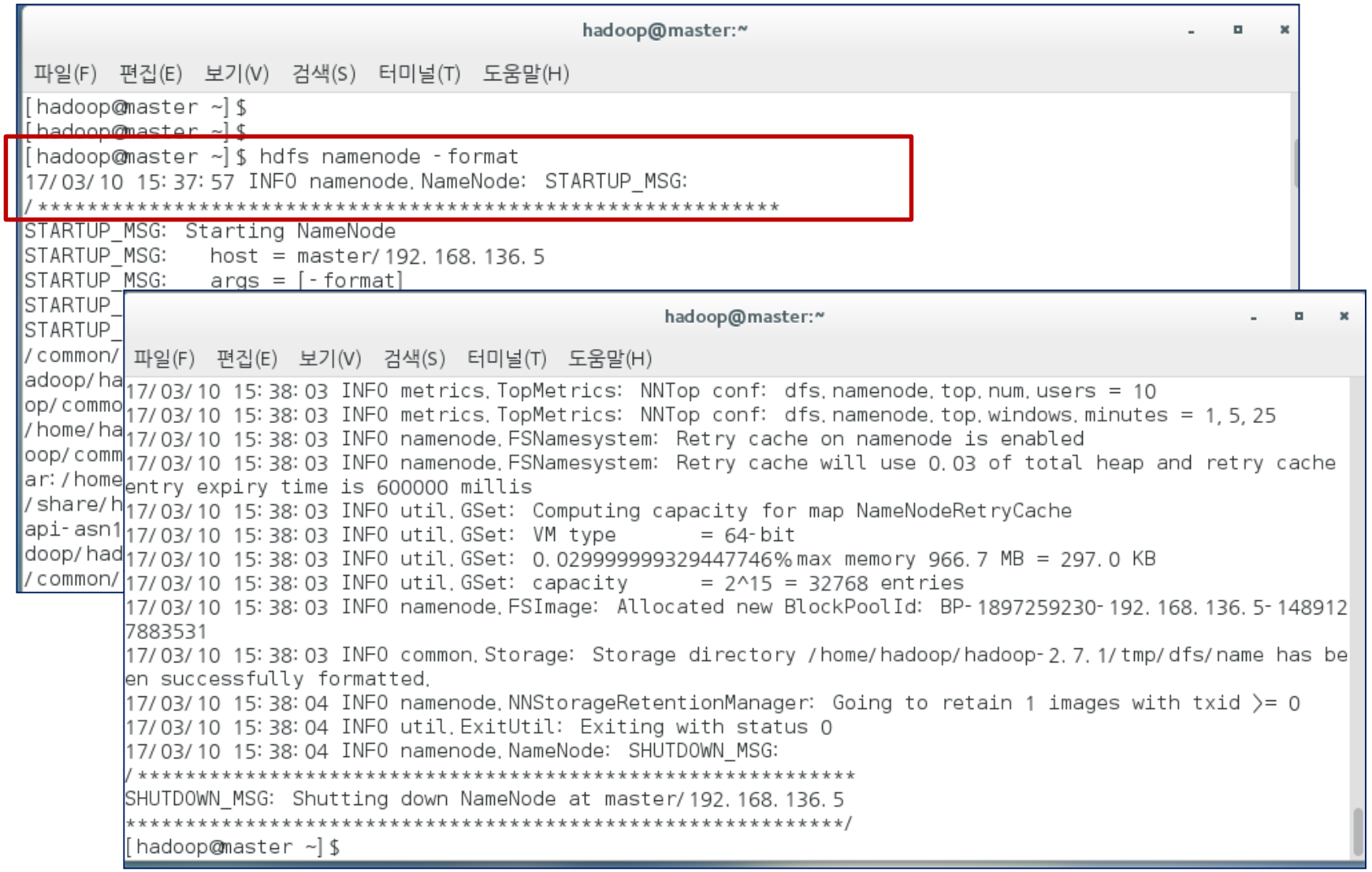
**2) Hadoop** 환경 변수(계정 **profile** 배포)

[hadoop@master ~]$ scp /home/hadoop/.bash_profile hadoop@slave1:~

[hadoop@master ~]$ scp /home/hadoop/.bash_profile hadoop@slave2:~

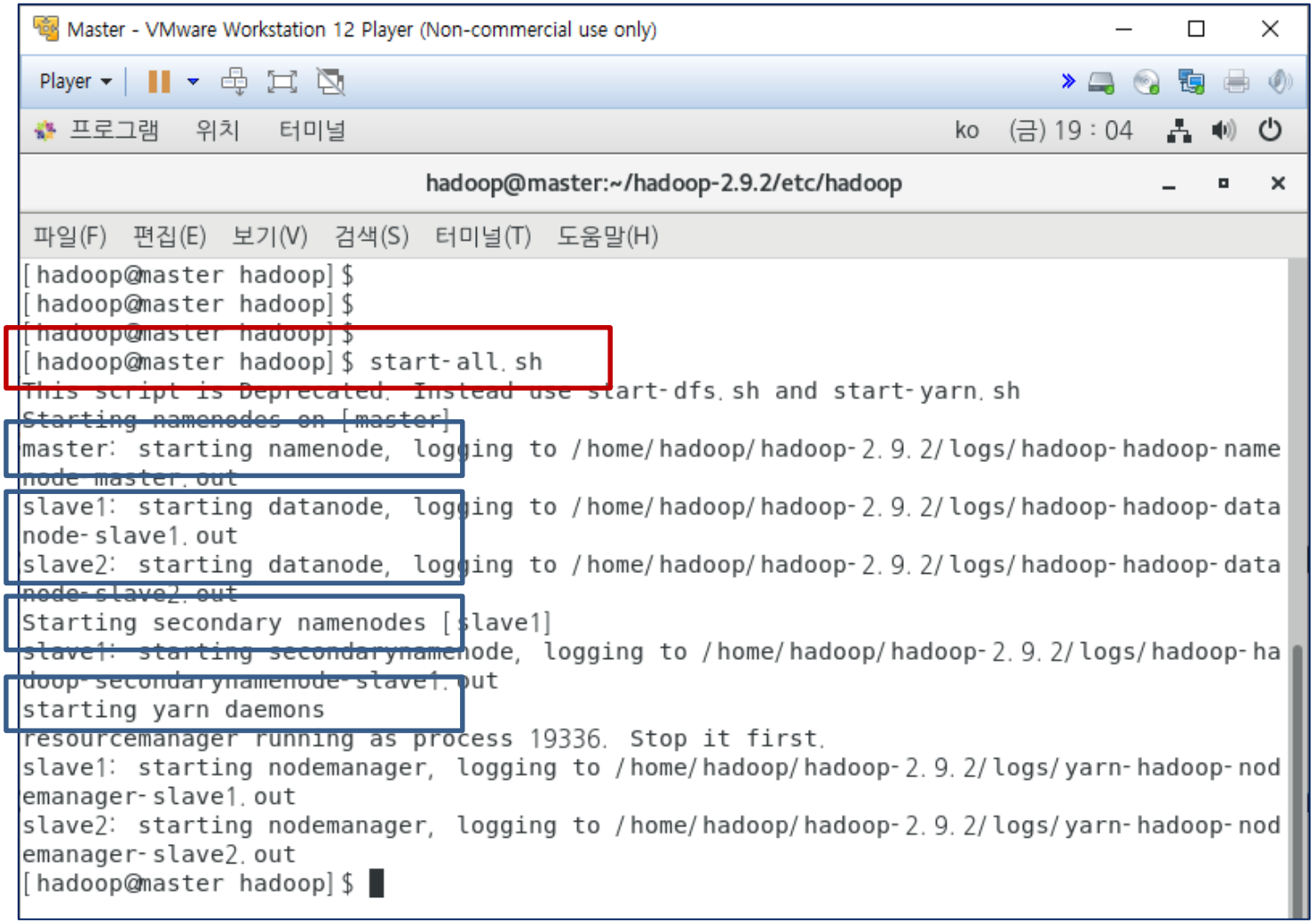


# HDFS 포멧(최초 사용시)

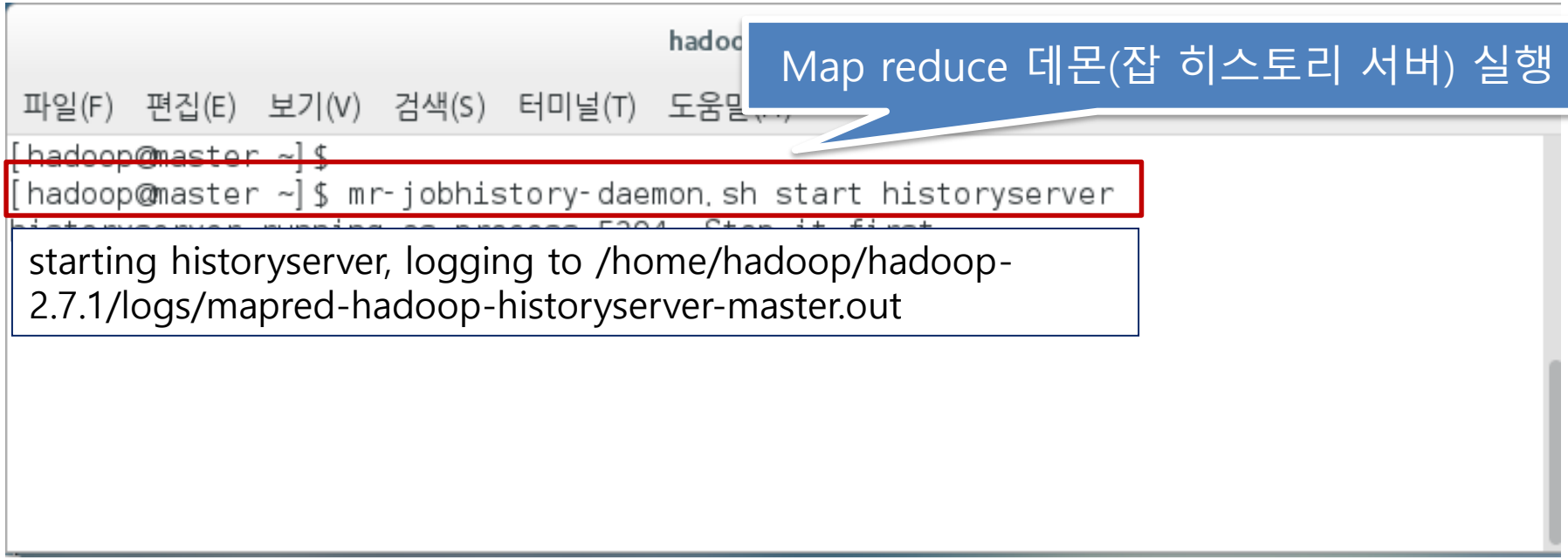


# 5. Hadoop/Yarn/Historyserver 시작

## 1) Hadoop/Yarn 시작

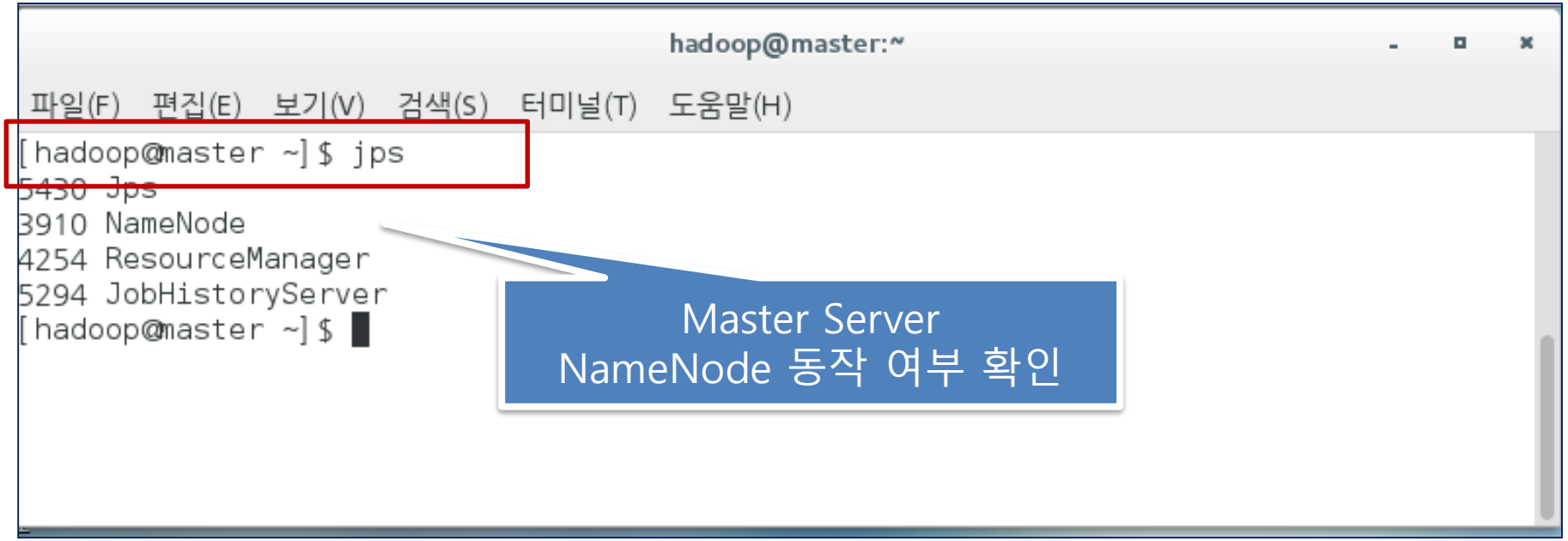


# 3) Historyserver 데몬 실행

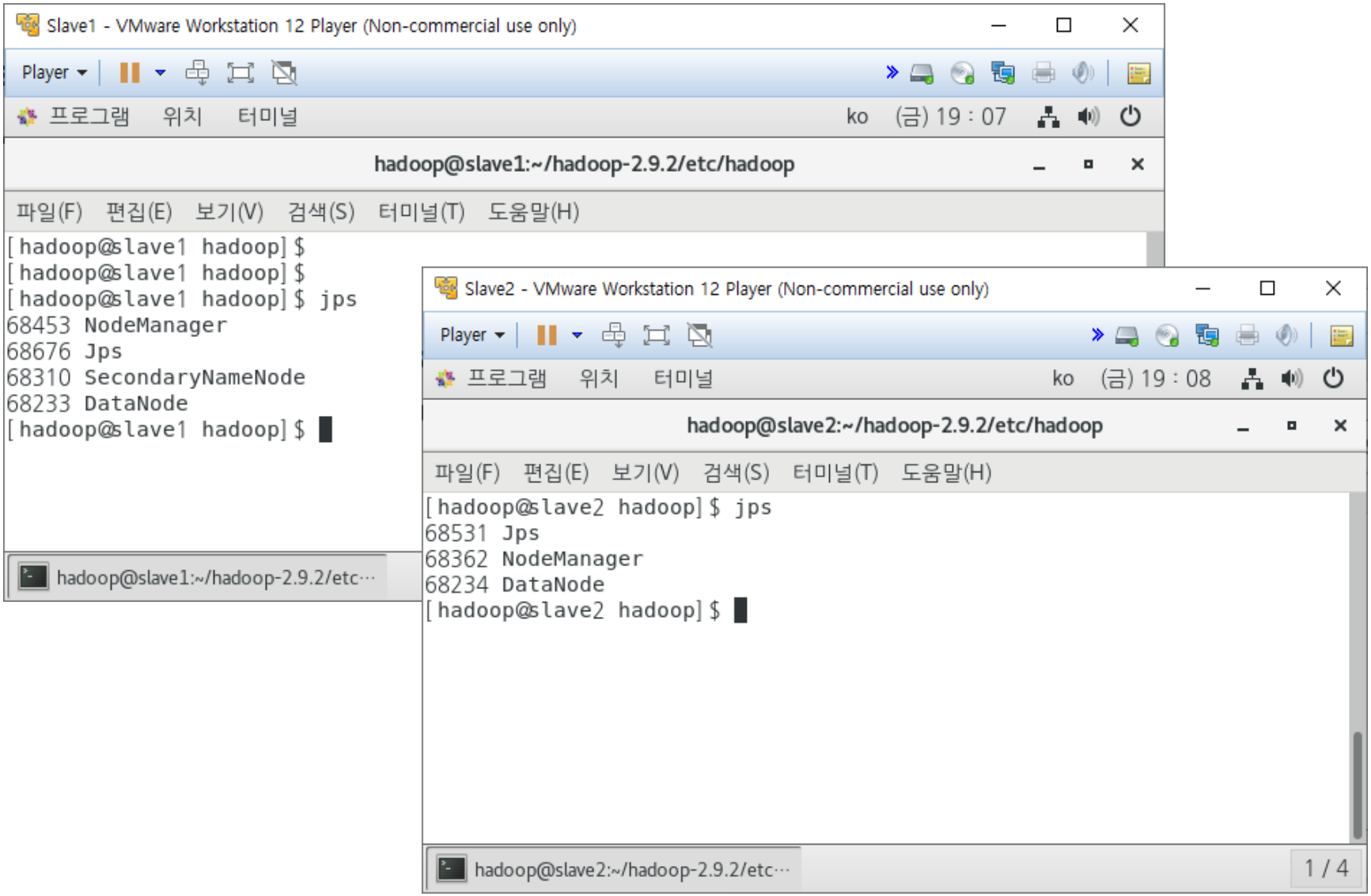


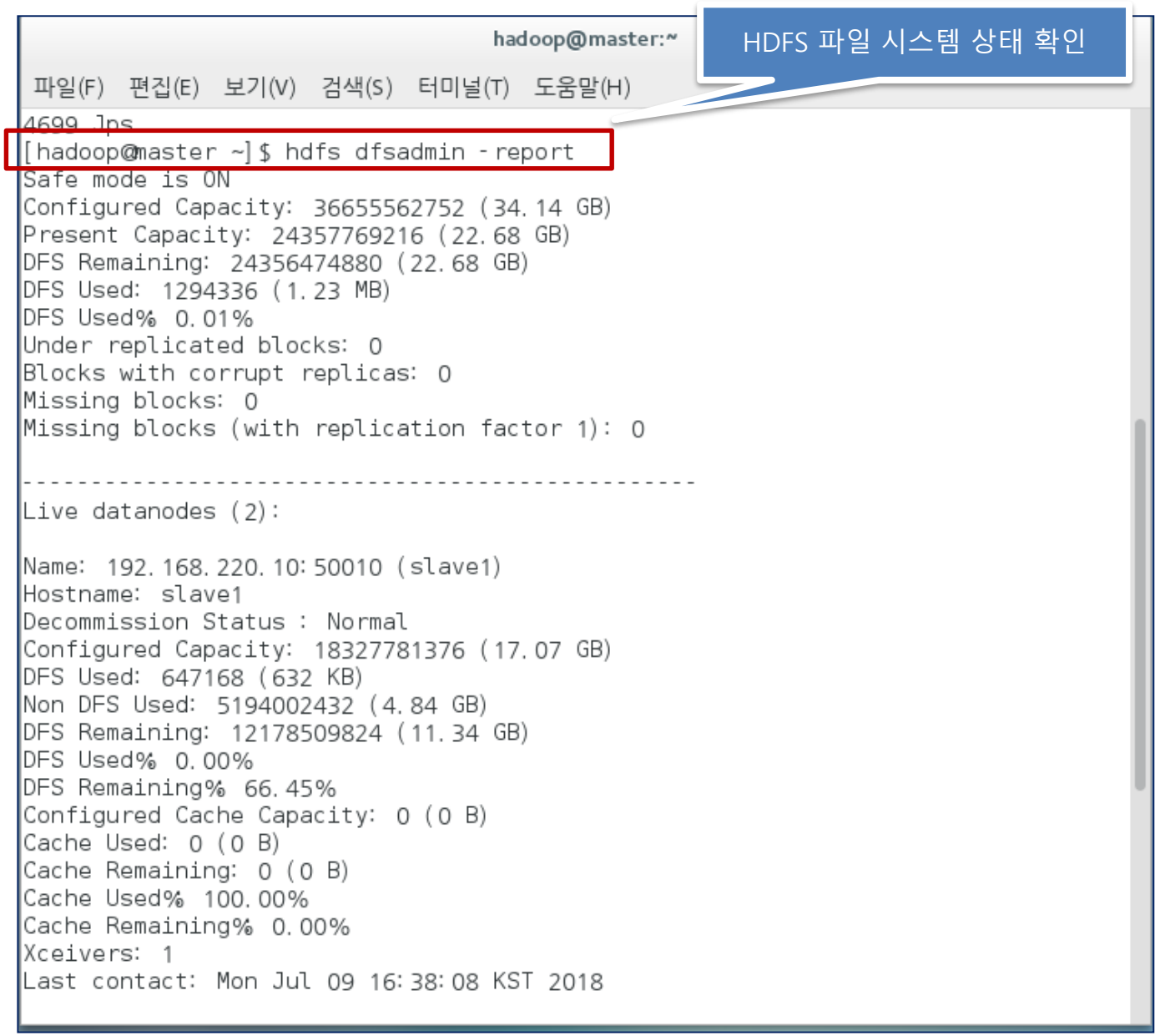
# 6. Hadoop system 상태 확인

❖ Name node용 서버에서 jsp 실행하면 NameNode, JobTracker 출력

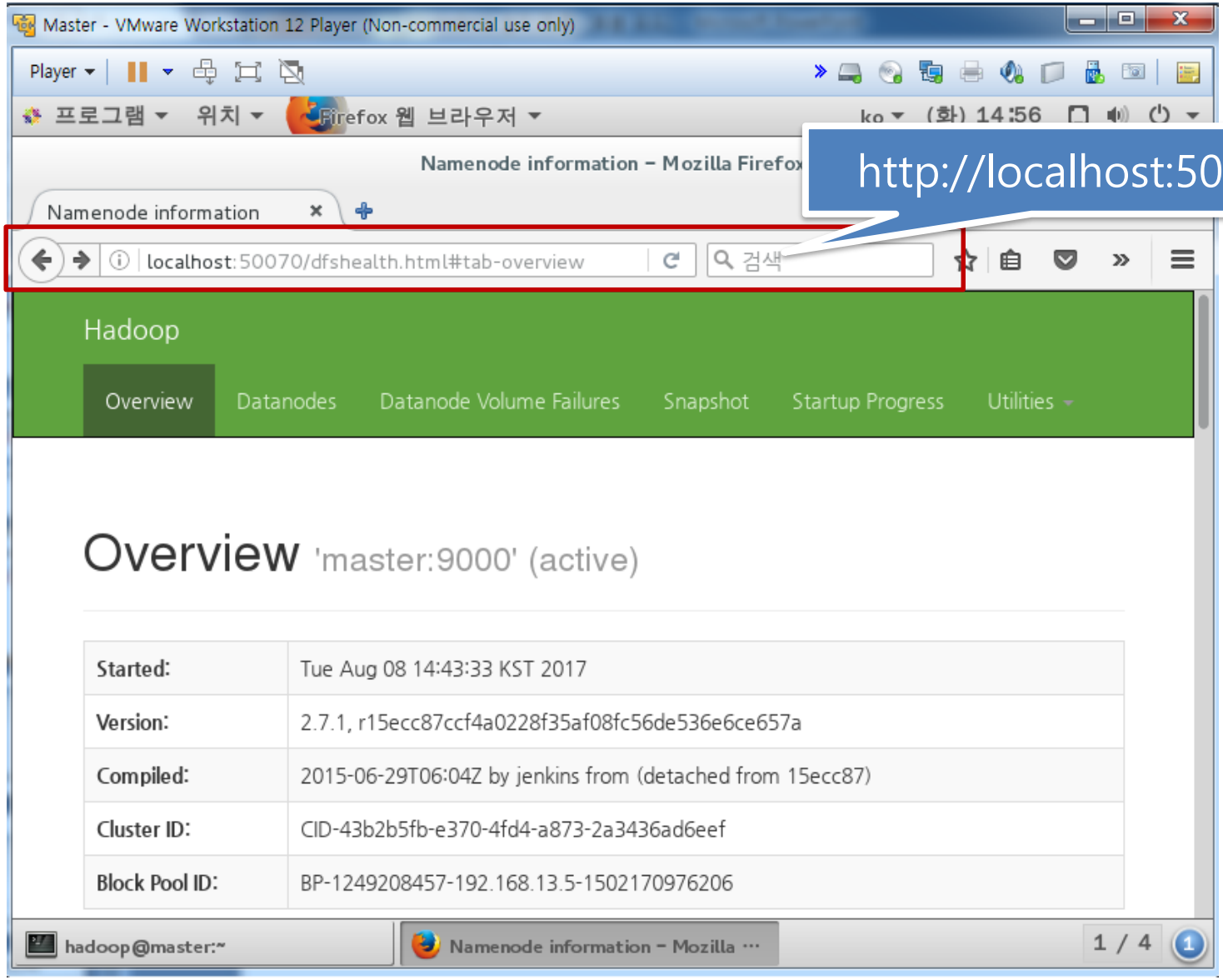


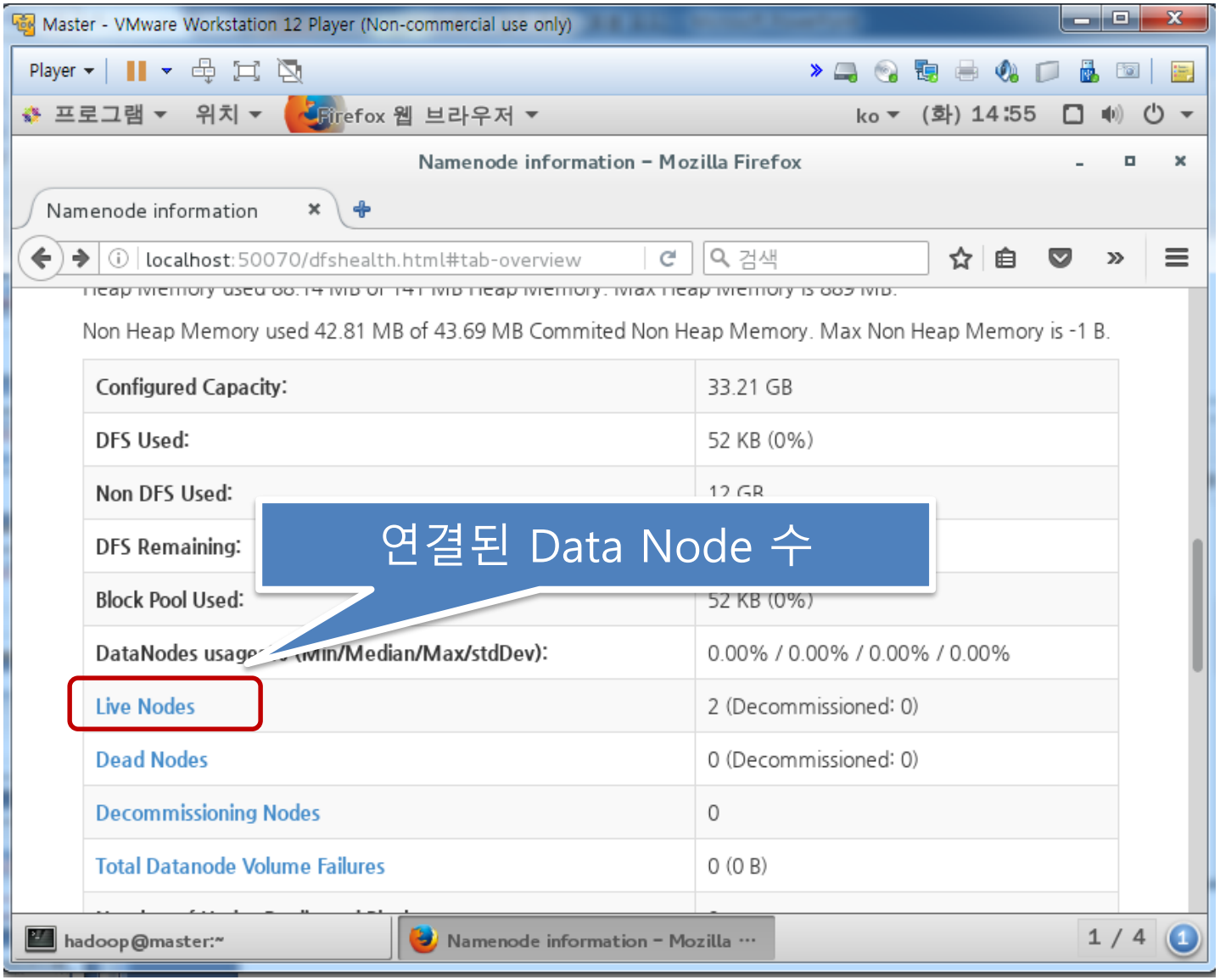
❖ **Data Node용 서버에서** jsp 실행하면 DataNode, SecondaryNameNode 출력




hadoop@master:~
HDFS 파일 시스템 상태 확인
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
4699 Jps
[hadoop@master ~]$ hdfs dfsadmin -report
Safe mode is ON
Configured Capacity: 36655562752 (34.14 GB)
Present Capacity: 24357769216 (22.68 GB)
DFS Remaining: 24356474880 (22.68 GB)
DFS Used: 1294336 (1.23 MB)
DFS Used% 0.01%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
-------------------------------------------------
Live datanodes (2):
Name: 192.168.220.10:50010 (slave1)
Hostname: slave1
Decommission Status : Normal
Configured Capacity: 18327781376 (17.07 GB)
DFS Used: 647168 (632 KB)
Non DFS Used: 5194002432 (4.84 GB)
DFS Remaining: 12178509824 (11.34 GB)
DFS Used% 0.00%
DFS Remaining% 66.45%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used% 100.00%
Cache Remaining% 0.00%
Xceivers: 1
Last contact: Mon Jul 09 16:38:08 KST 2018

# Hadoop 상태 확인


Master - VMware Workstation 12 Player (Non-commercial use only)
Player
프로그램
위치
Firefox 웹 브라우저
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-overview
검색
http://localhost:50070
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'master:9000' (active)
Started: Tue Aug 08 14:43:33 KST 2017
Version: 2.7.1, r15ecc87ccf4a0228f35af08fc56de536e6ce657a
Compiled: 2015-06-29T06:04Z by jenkins from (detached from 15ecc87)
Cluster ID: CID-43b2b5fb-e370-4fd4-a873-2a3436ad6eef
Block Pool ID: BP-1249208457-192.168.13.5-1502170976206
hadoop@master:~
Namenode information - Mozilla ...
1 / 4


Master - VMware Workstation 12 Player (Non-commercial use only)
Player
프로그램
위치
Firefox 웹 브라우저
ko
(화) 14:55
Namenode information – Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-overview
검색
Non Heap Memory used 42.81 MB of 43.69 MB Commited Non Heap Memory. Max Non Heap Memory is -1 B.
Configured Capacity: 33.21 GB
DFS Used: 52 KB (0%)
Non DFS Used:
DFS Remaining:
Block Pool Used: 52 KB (0%)
DataNodes usage (Min/Median/Max/stdDev): 0.00% / 0.00% / 0.00% / 0.00%
Live Nodes 2 (Decommissioned: 0)
Dead Nodes 0 (Decommissioned: 0)
Decommissioning Nodes 0
Total Datanode Volume Failures 0 (0 B)
연결된 Data Node 수
hadoop@master:~
Namenode information – Mozilla ...
1 / 4

# 7. Hadoop system 종료