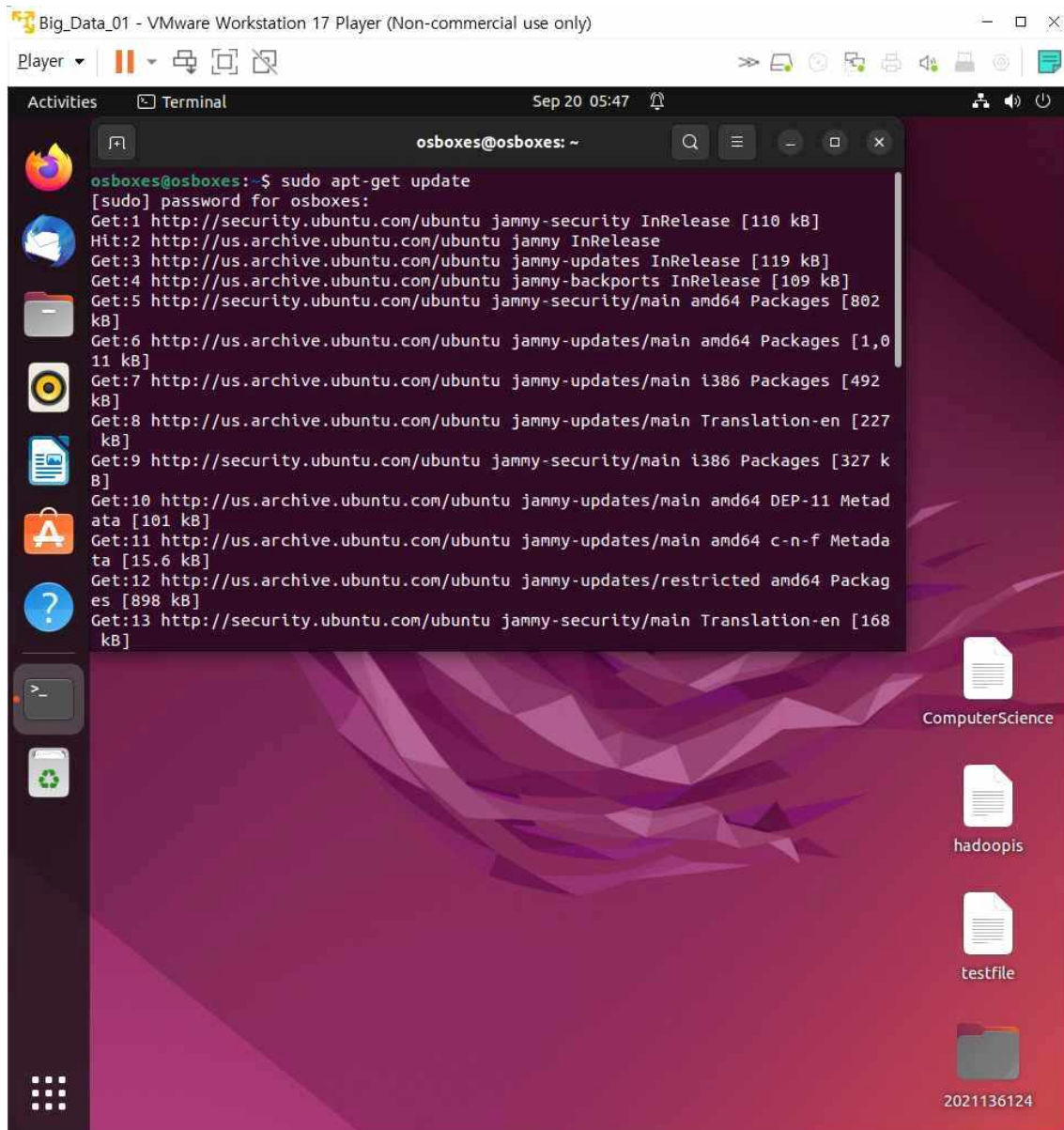
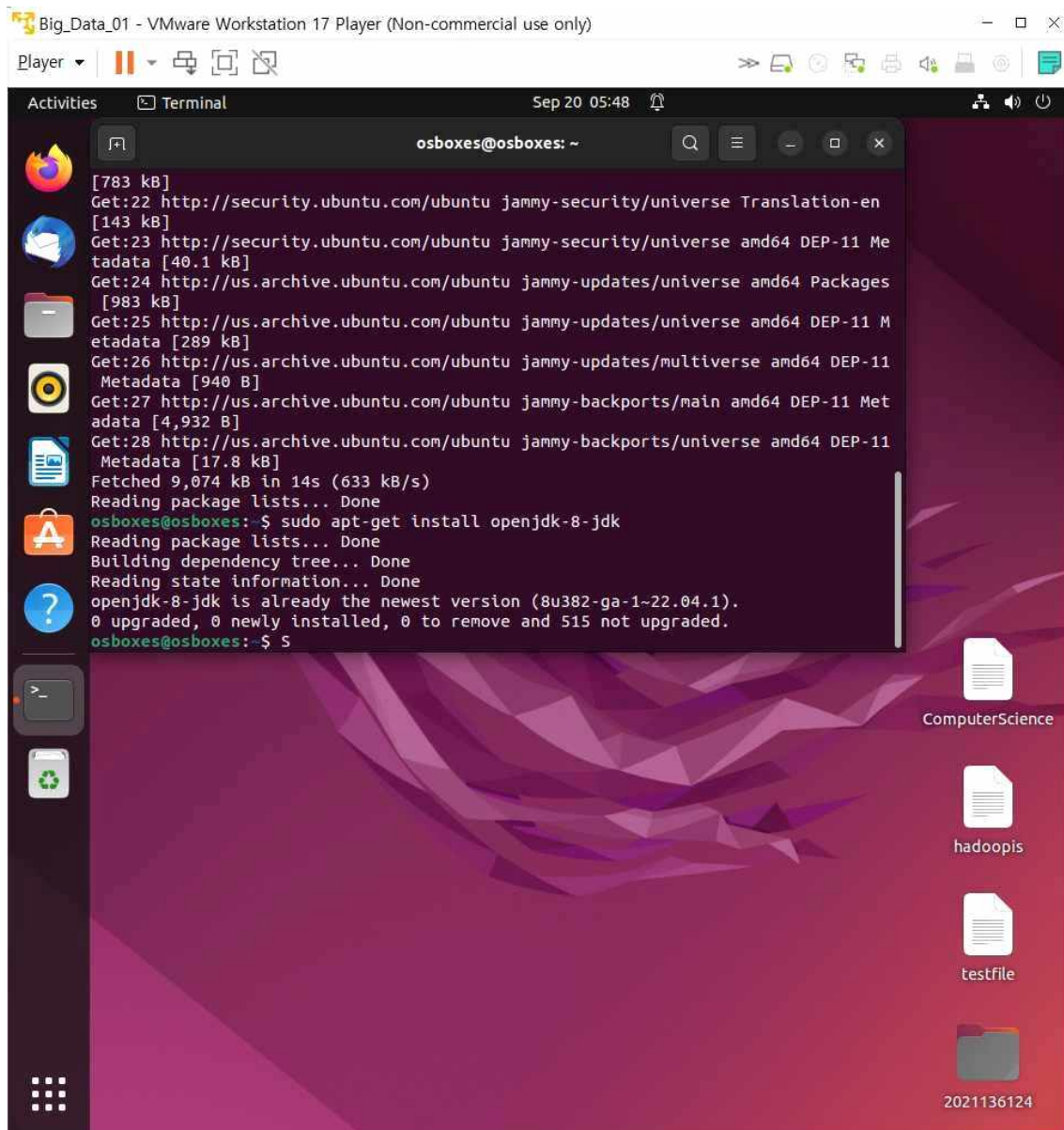
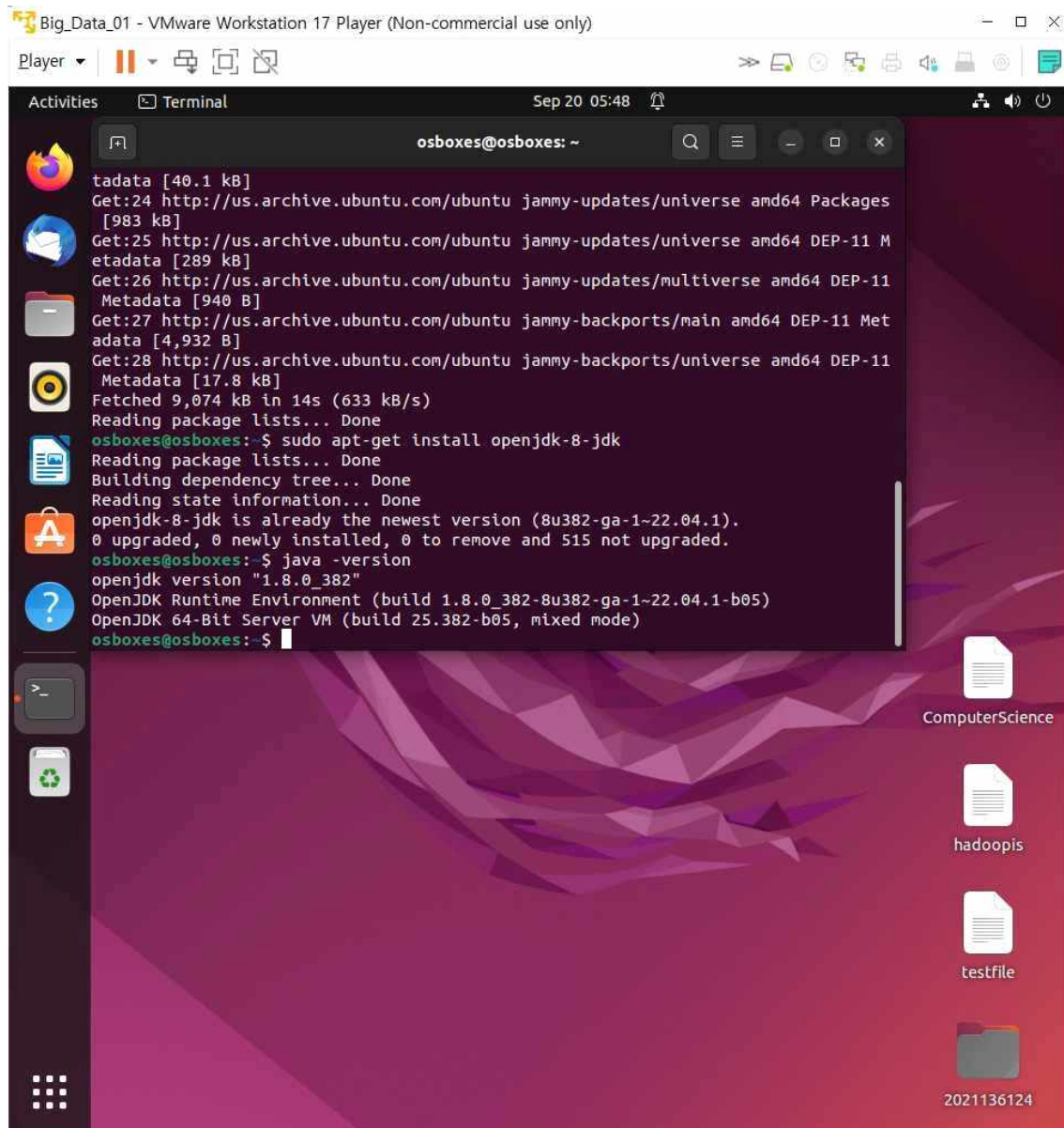


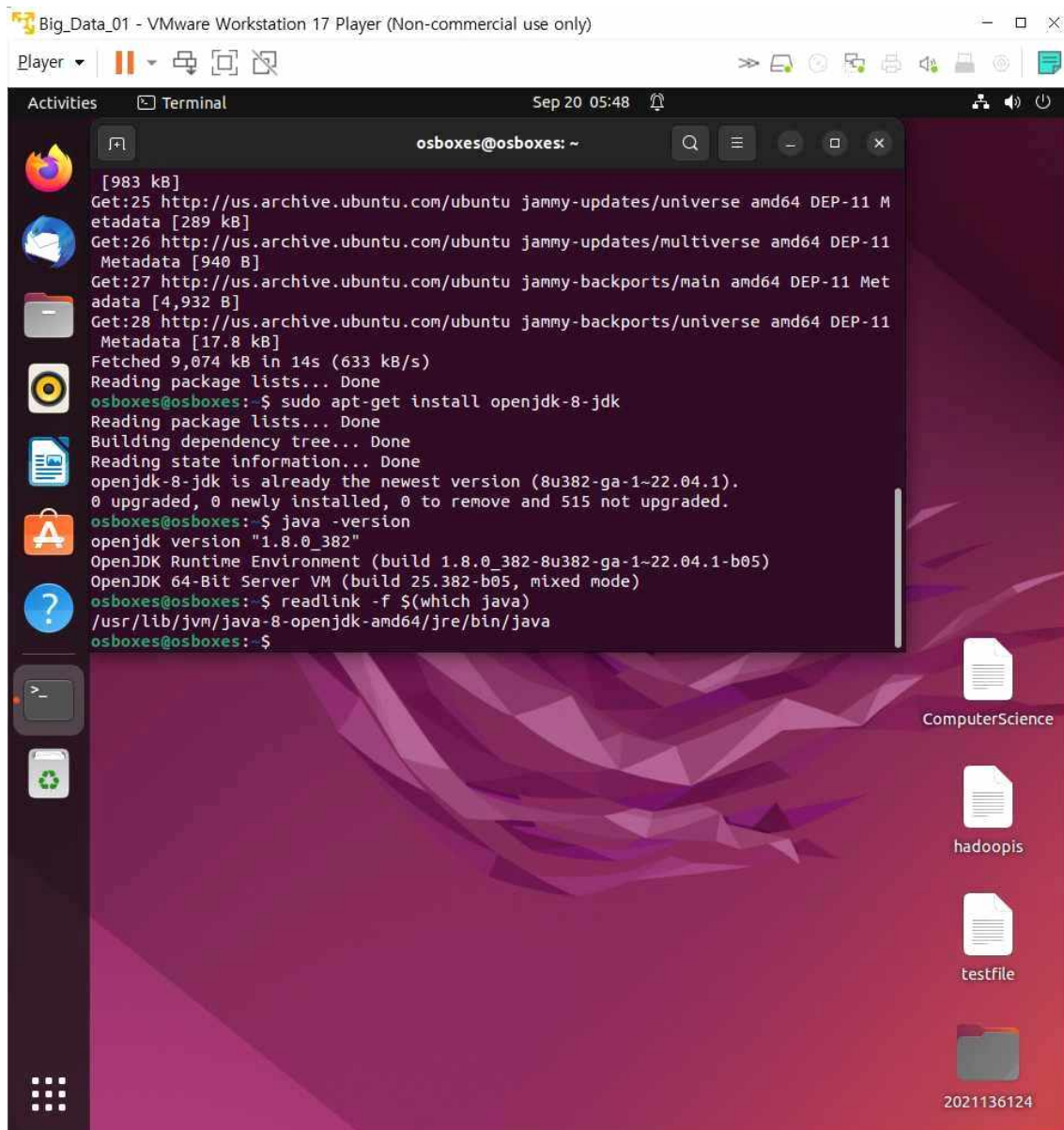
## java 설치



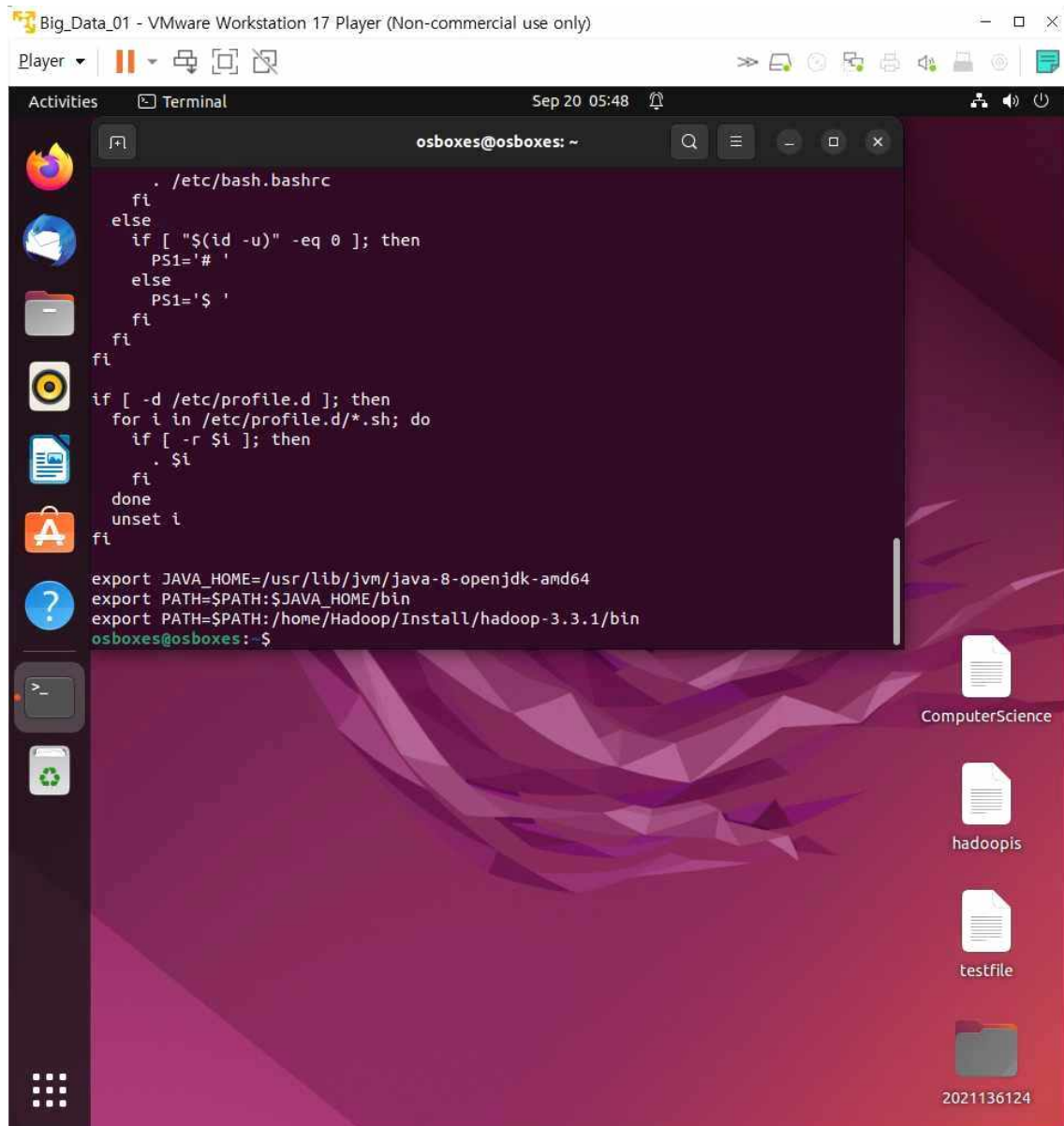




JAVA\_HOME 설정

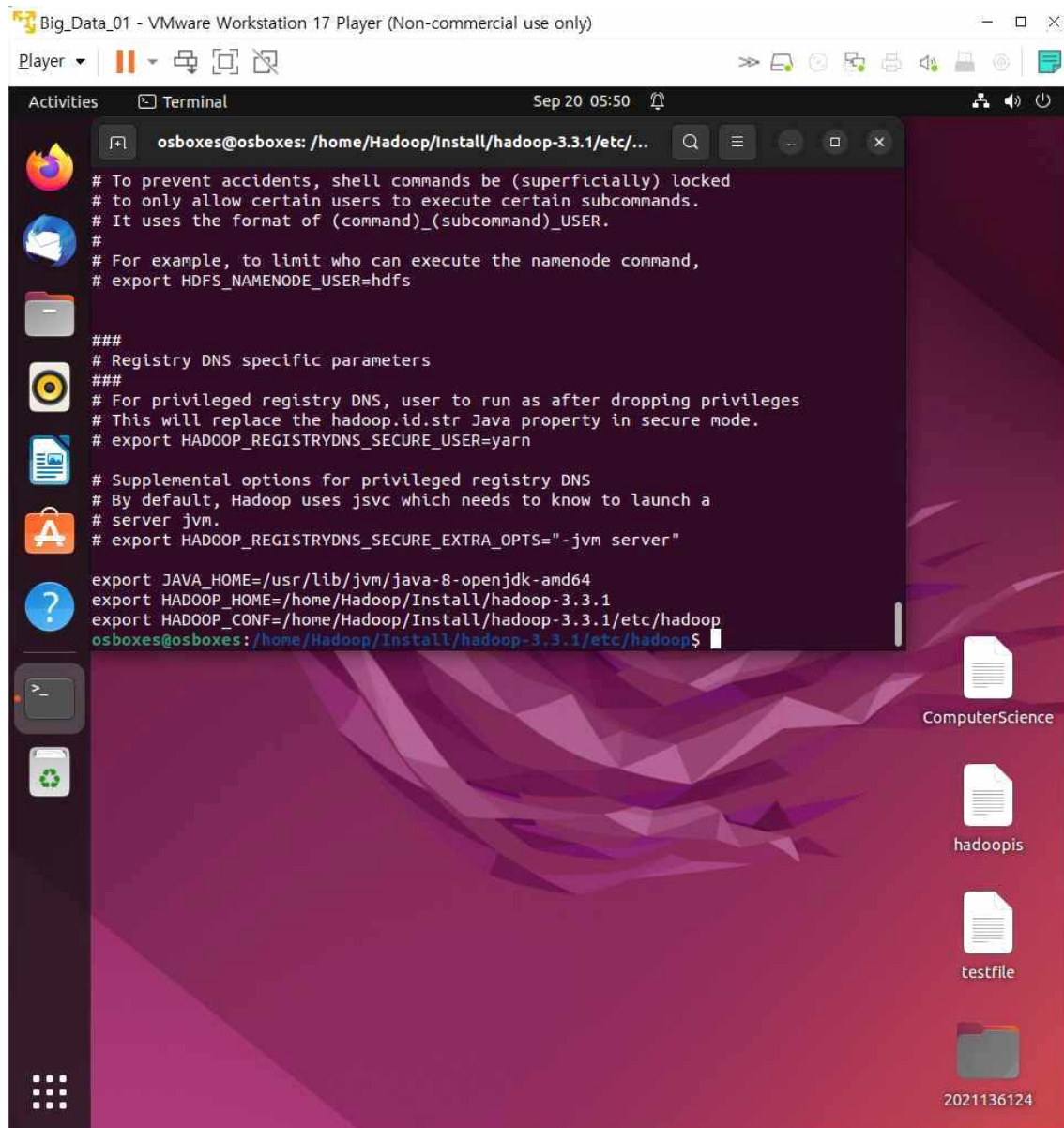




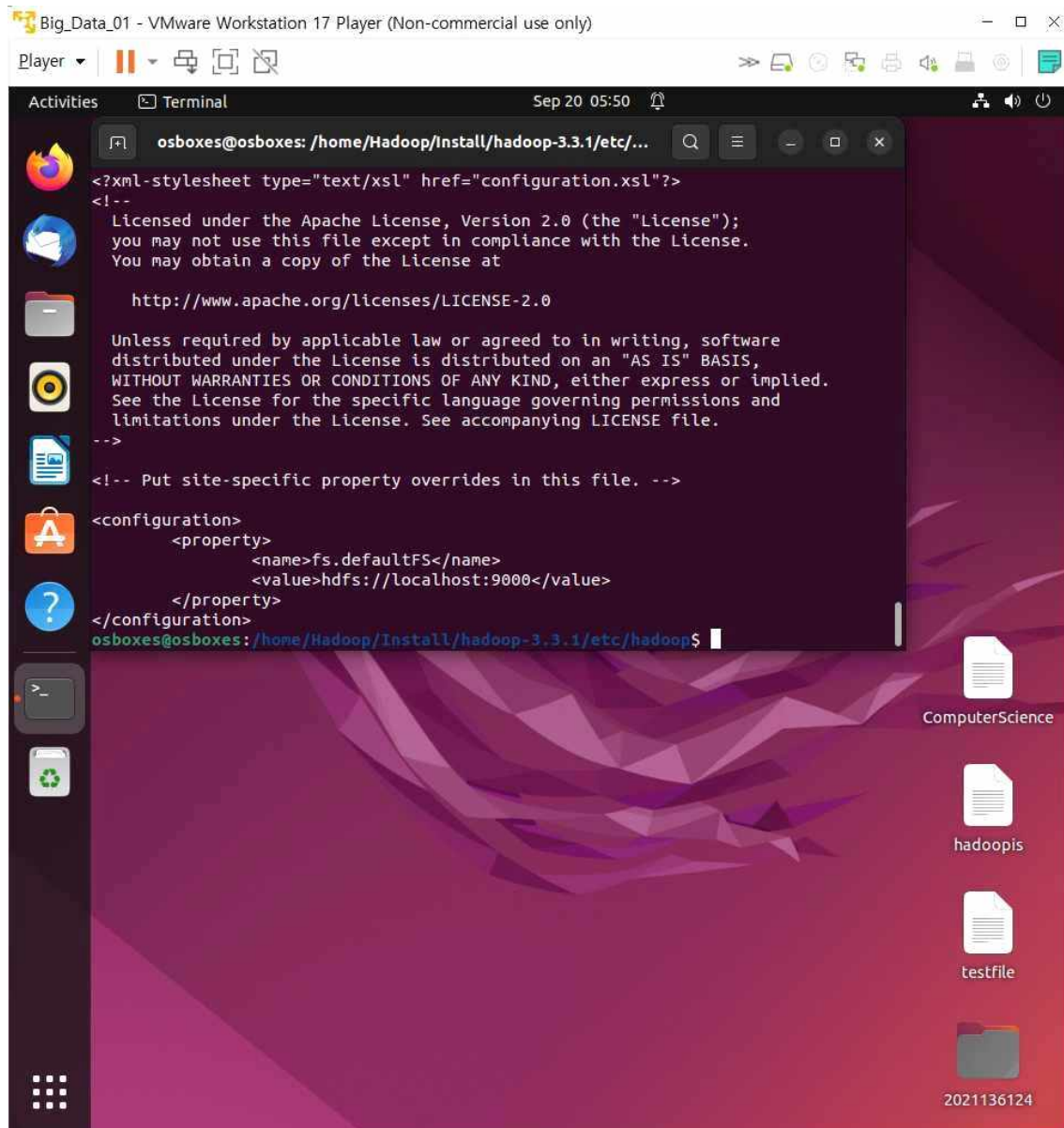


hadoop 설치

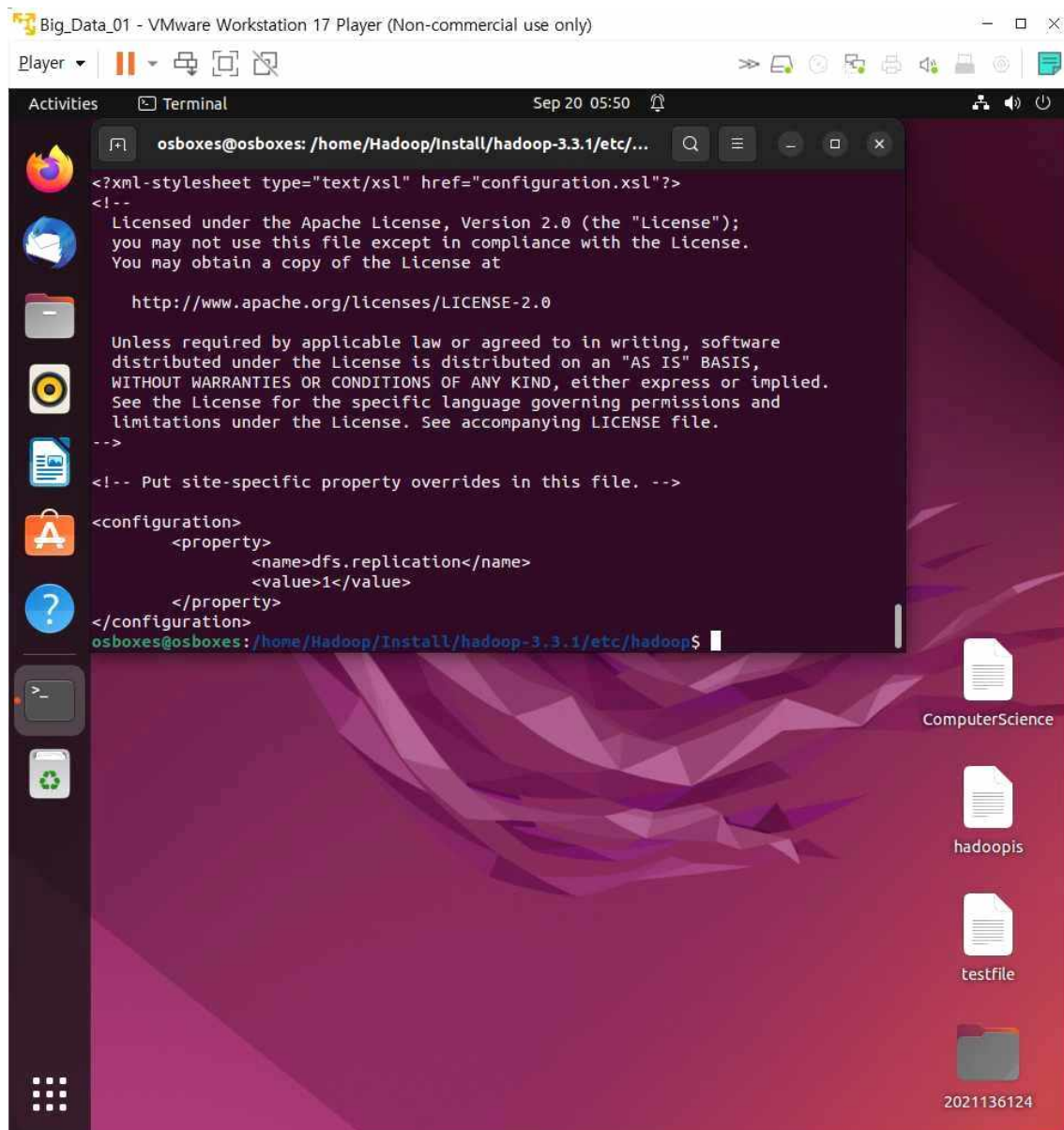
hadoop 설정 파일 (hadoop-env.sh) 수정



core-site.xml

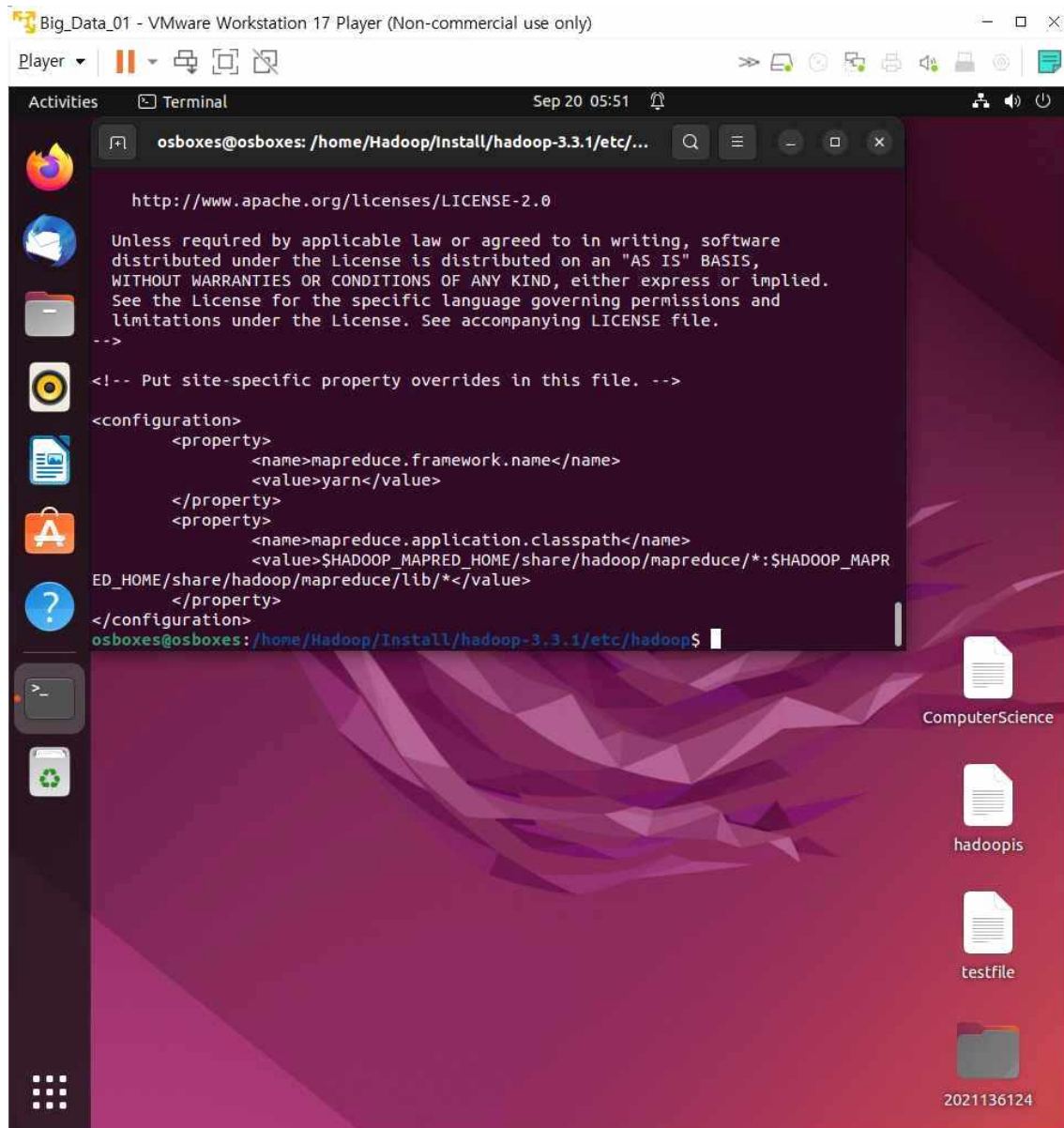


hdfs-site.xml

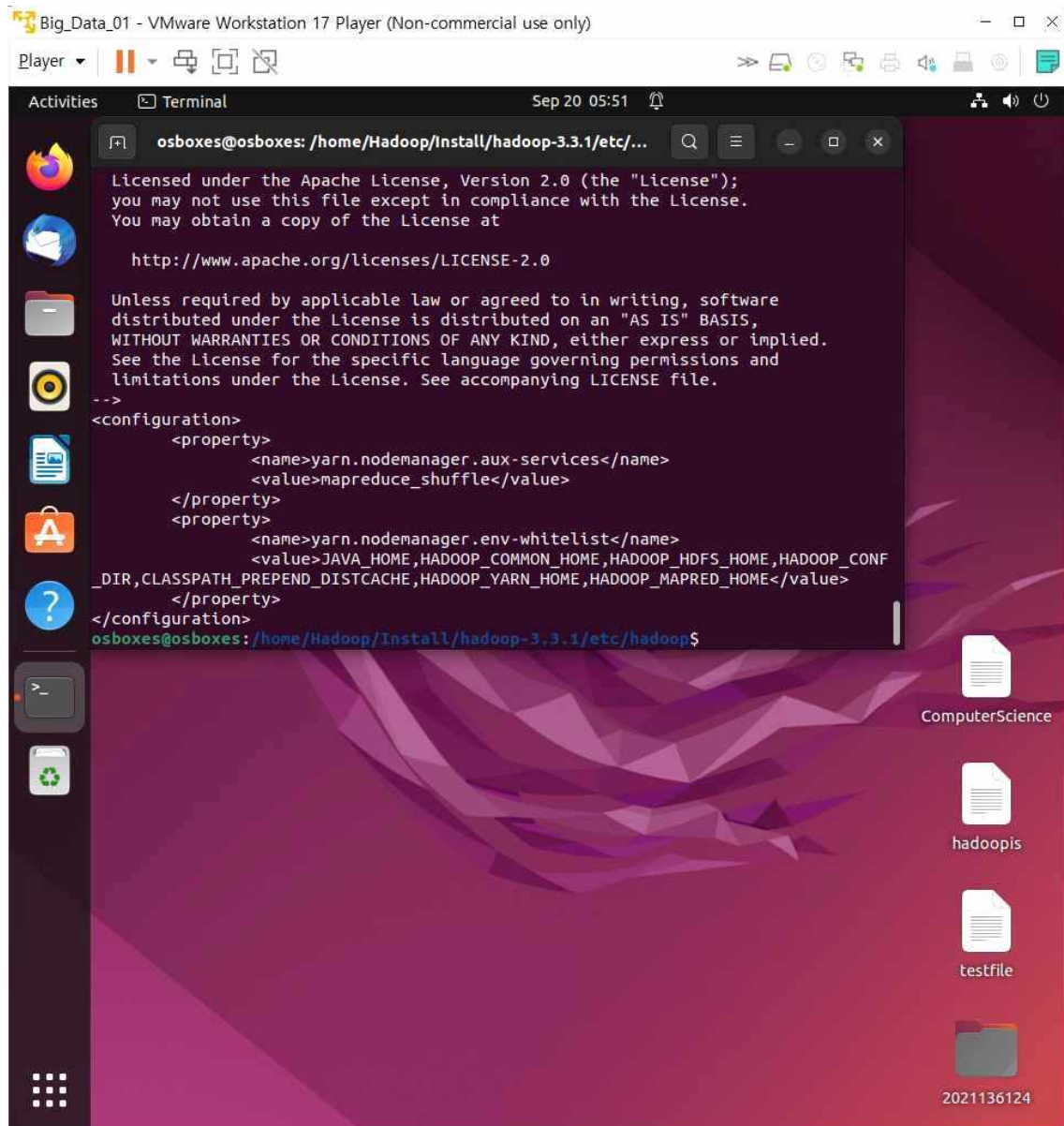


mapred-site.xml

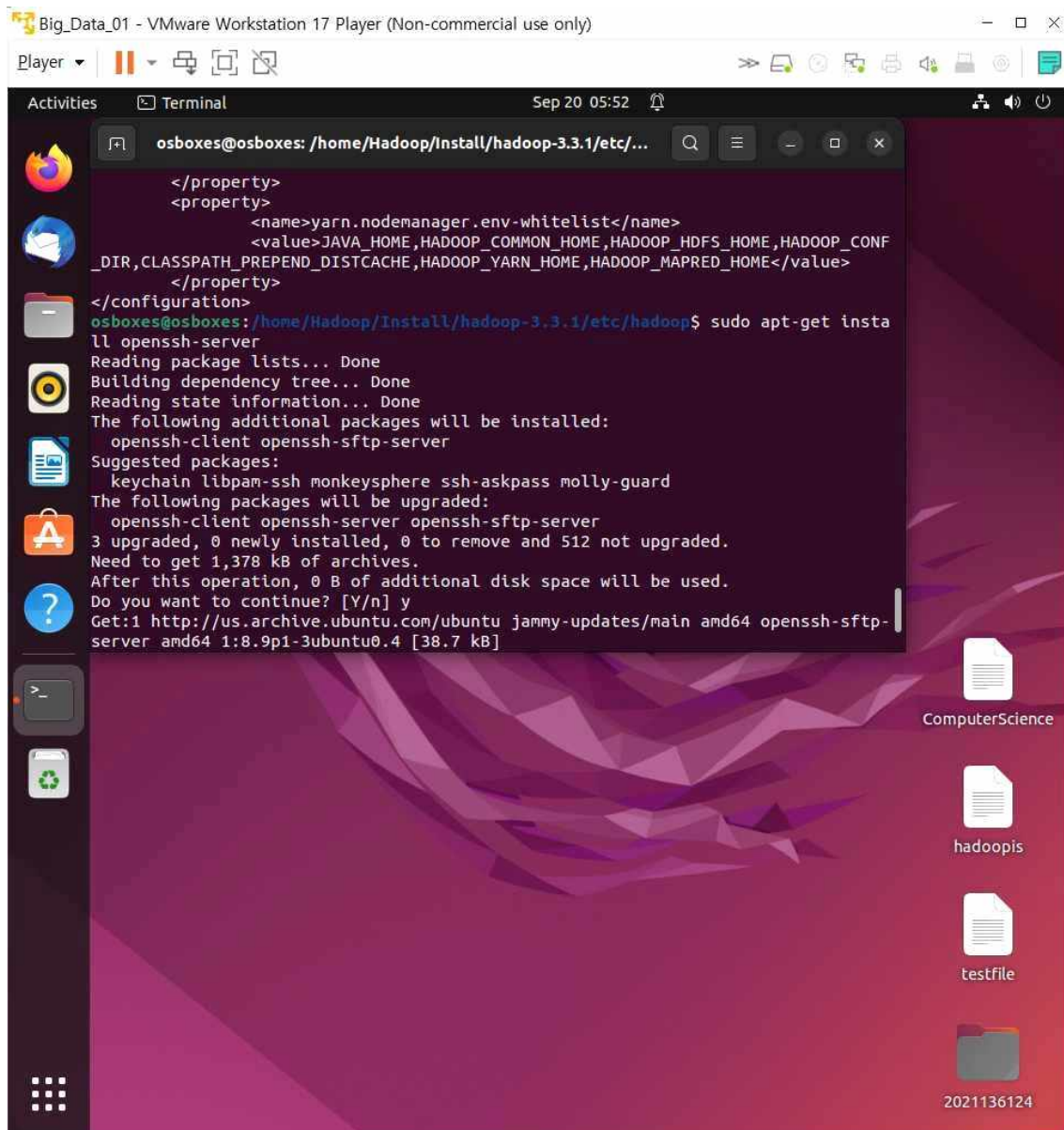


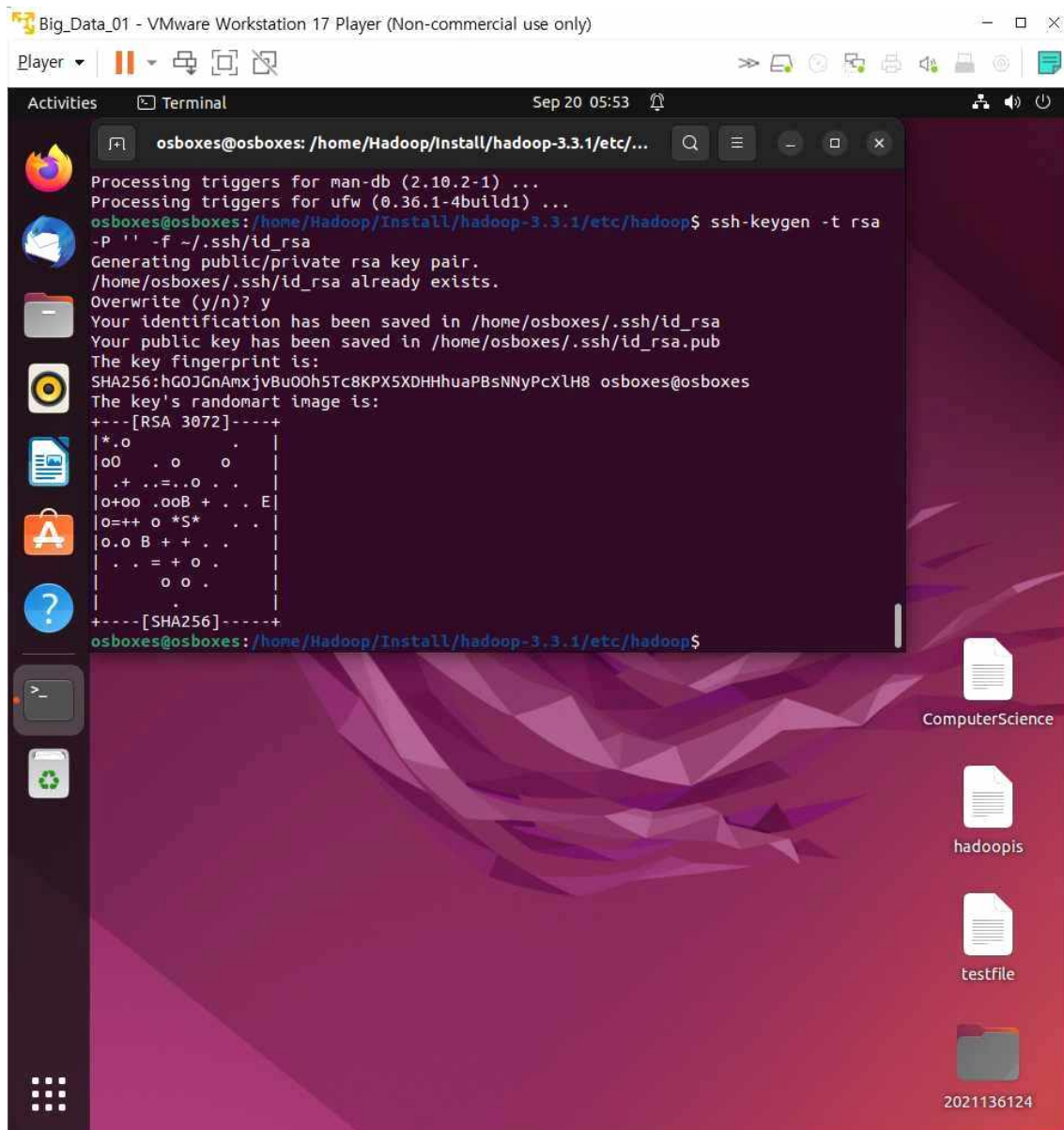


y a r n - s i t e . x m l

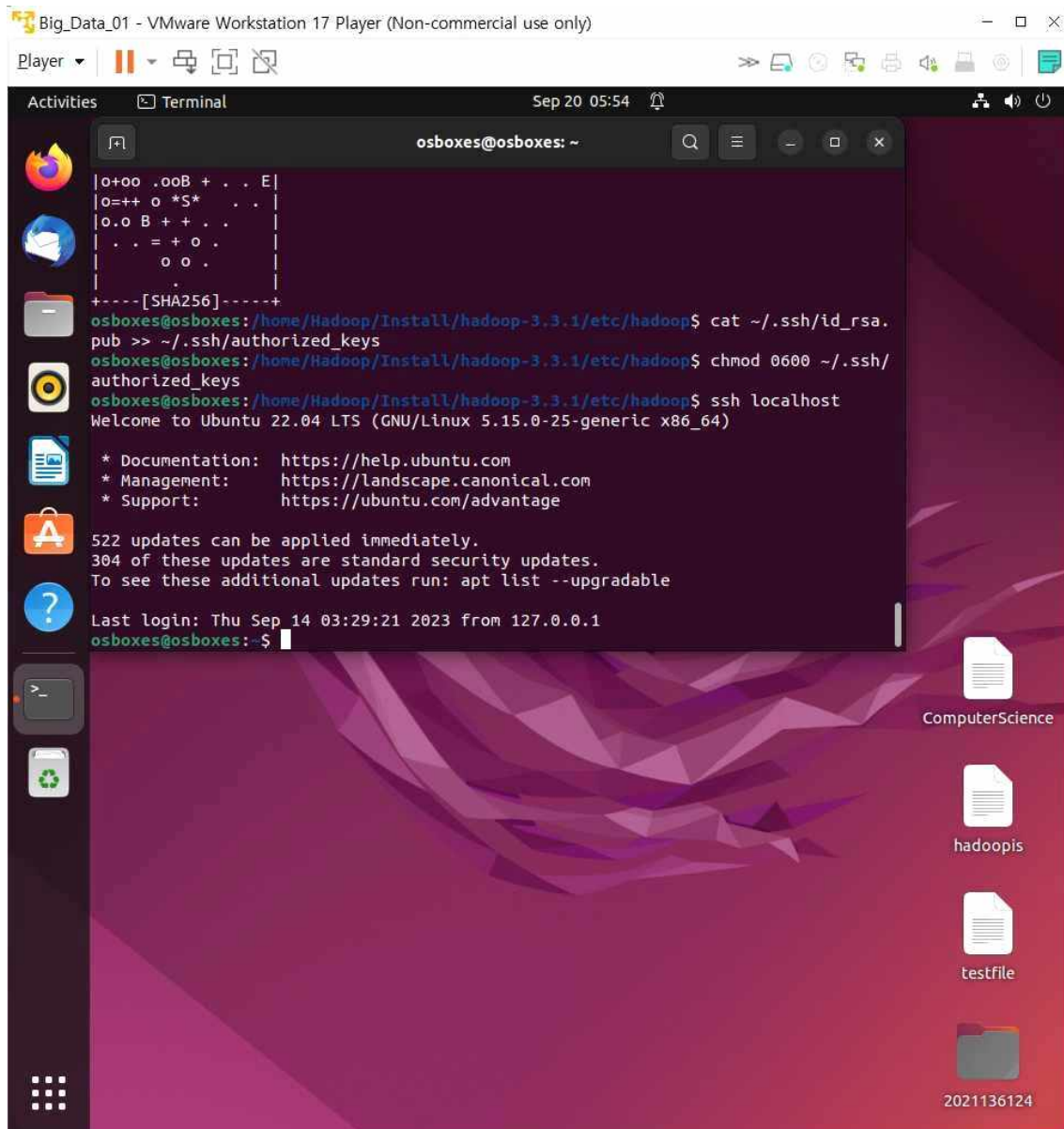


ssh 설정

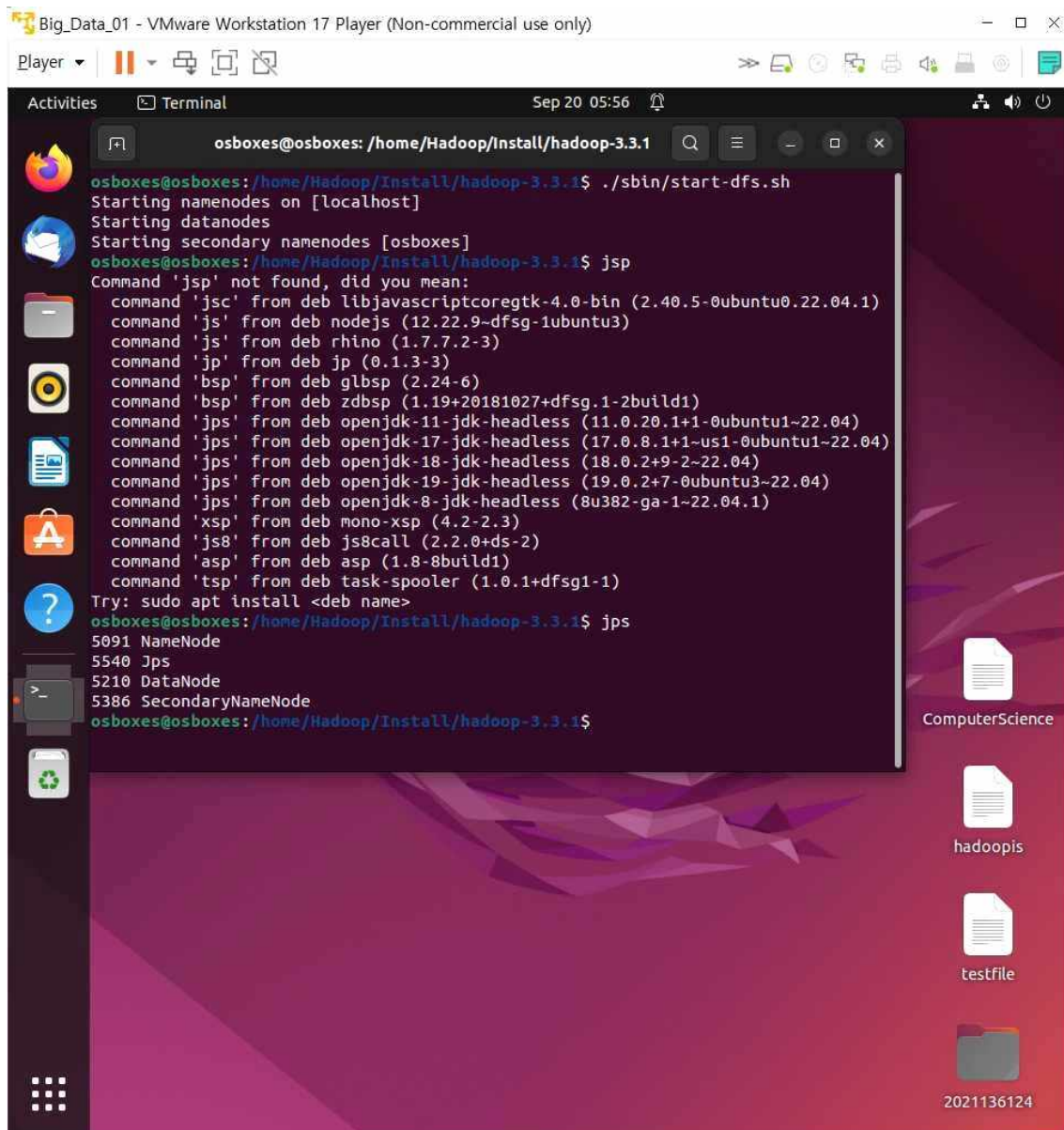


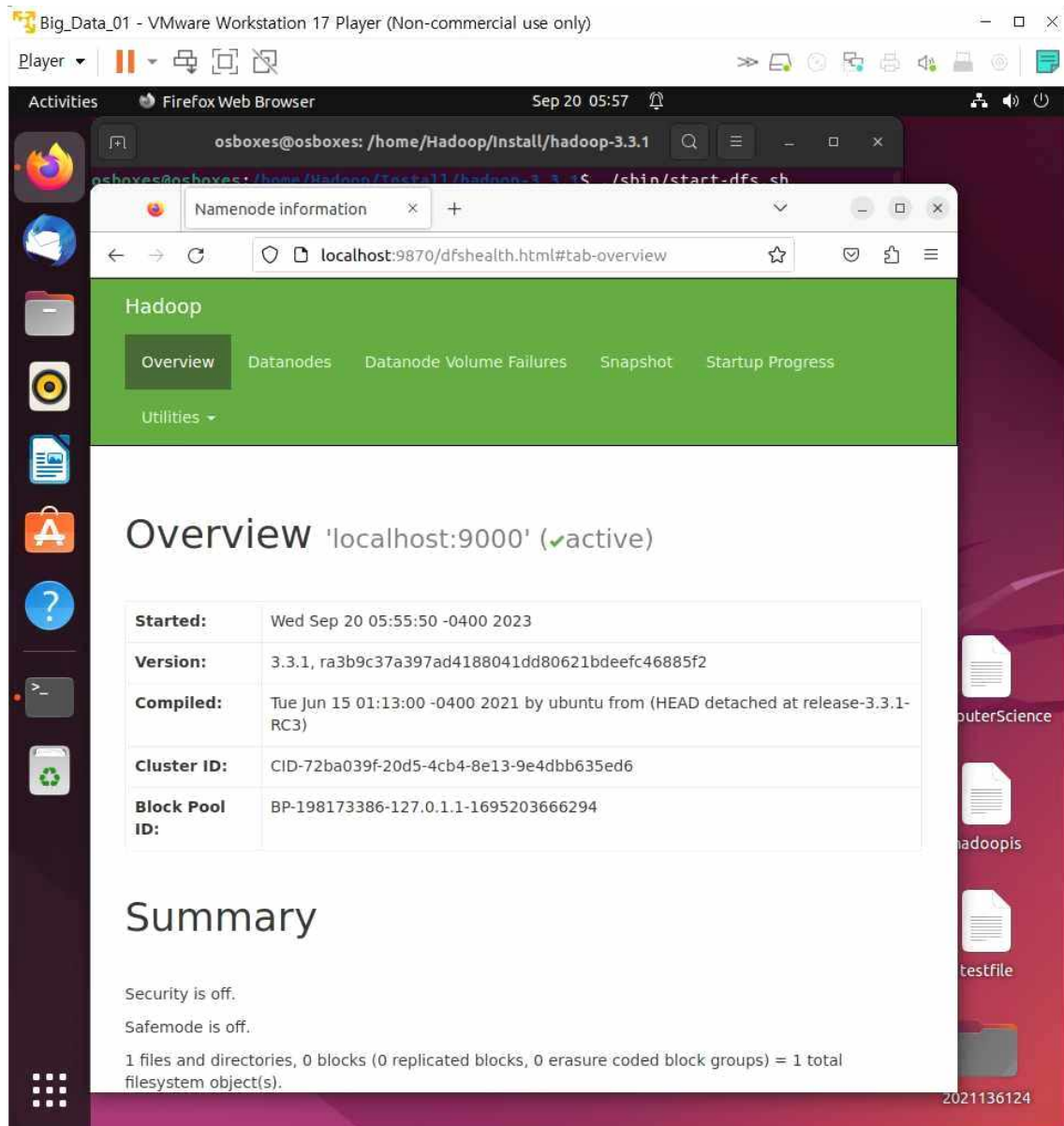






HDFS 데몬 실행





TARN 데몬 실행

Big\_Data\_01 - VMware Workstation 17 Player (Non-commercial use only)

Player

Sep 20 05:57

Activities Terminal

osboxes@osboxes: /home/Hadoop/Install/hadoop-3.3.1

```
command 'js' from deb rhino (1.7.7.2-3)
command 'jp' from deb jp (0.1.3-3)
command 'bsp' from deb glbsp (2.24-6)
command 'bsp' from deb zdbsp (1.19+20181027+dfsg.1-2build1)
command 'jps' from deb openjdk-11-jdk-headless (11.0.20.1+1-0ubuntu1~22.04)
command 'jps' from deb openjdk-17-jdk-headless (17.0.8.1+1-us1-0ubuntu1~22.04)
command 'jps' from deb openjdk-18-jdk-headless (18.0.2+9-2~22.04)
command 'jps' from deb openjdk-19-jdk-headless (19.0.2+7-0ubuntu3~22.04)
command 'jps' from deb openjdk-8-jdk-headless (8u382-ga-1-22.04.1)
command 'xsp' from deb mono-xsp (4.2-2.3)
command 'js8' from deb js8call (2.2.0+ds-2)
command 'asp' from deb asp (1.8-8build1)
command 'tsp' from deb task-spooler (1.0.1+dfsg1-1)
Try: sudo apt install <deb name>
osboxes@osboxes:/home/Hadoop/Install/hadoop-3.3.1$ jps
5091 NameNode
5540 Jps
5210 DataNode
5386 SecondaryNameNode
osboxes@osboxes:/home/Hadoop/Install/hadoop-3.3.1$ ./sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
osboxes@osboxes:/home/Hadoop/Install/hadoop-3.3.1$ jps
5091 NameNode
6276 Jps
6024 ResourceManager
6137 NodeManager
5210 DataNode
5386 SecondaryNameNode
osboxes@osboxes:/home/Hadoop/Install/hadoop-3.3.1$
```

Block Pool ID:	BP-198173386-127.0.1.1-1695203666294
----------------	--------------------------------------

## Summary

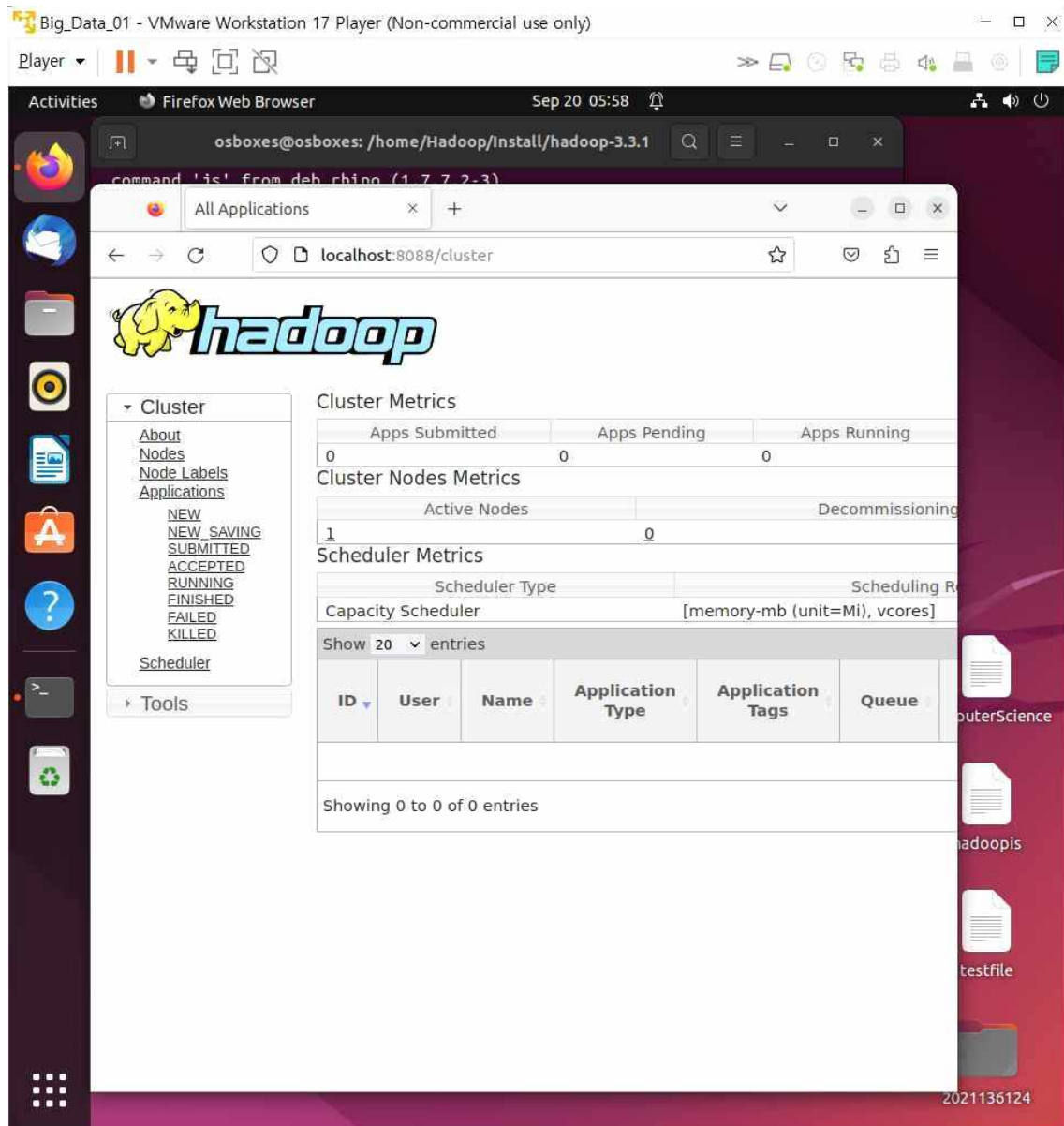
Security is off.

Safemode is off.

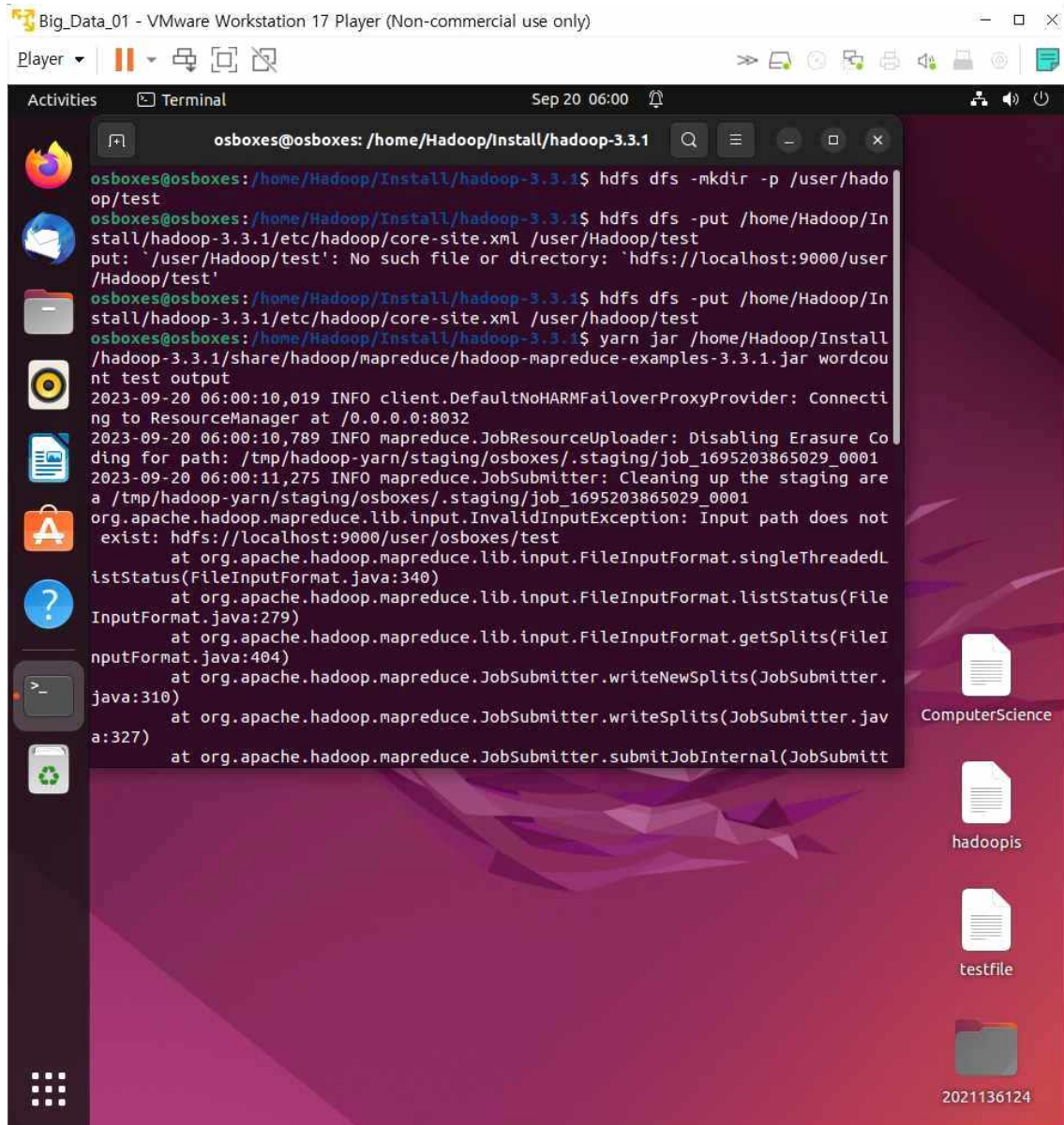
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

2021136124



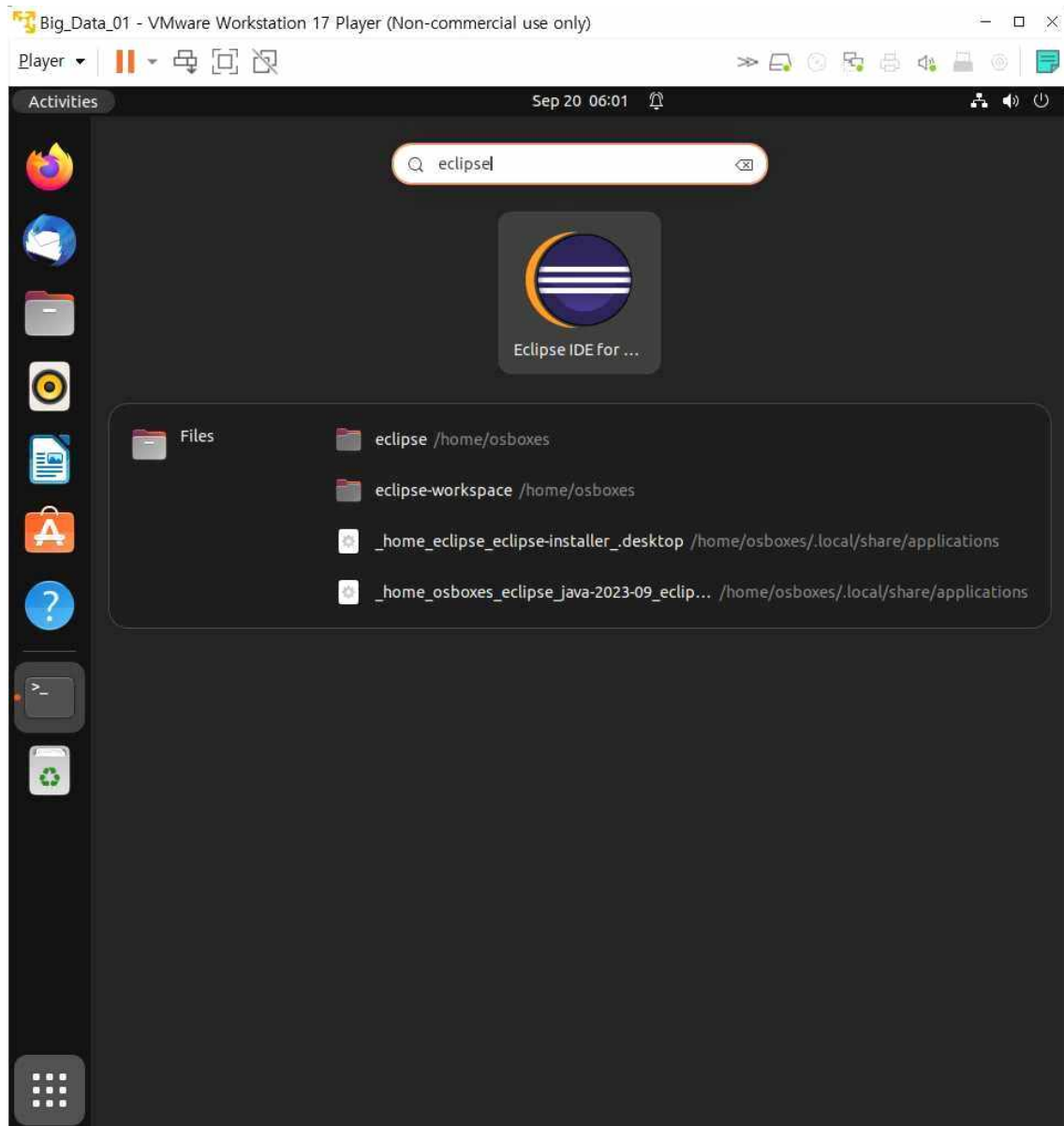


예제 프로그램 실행



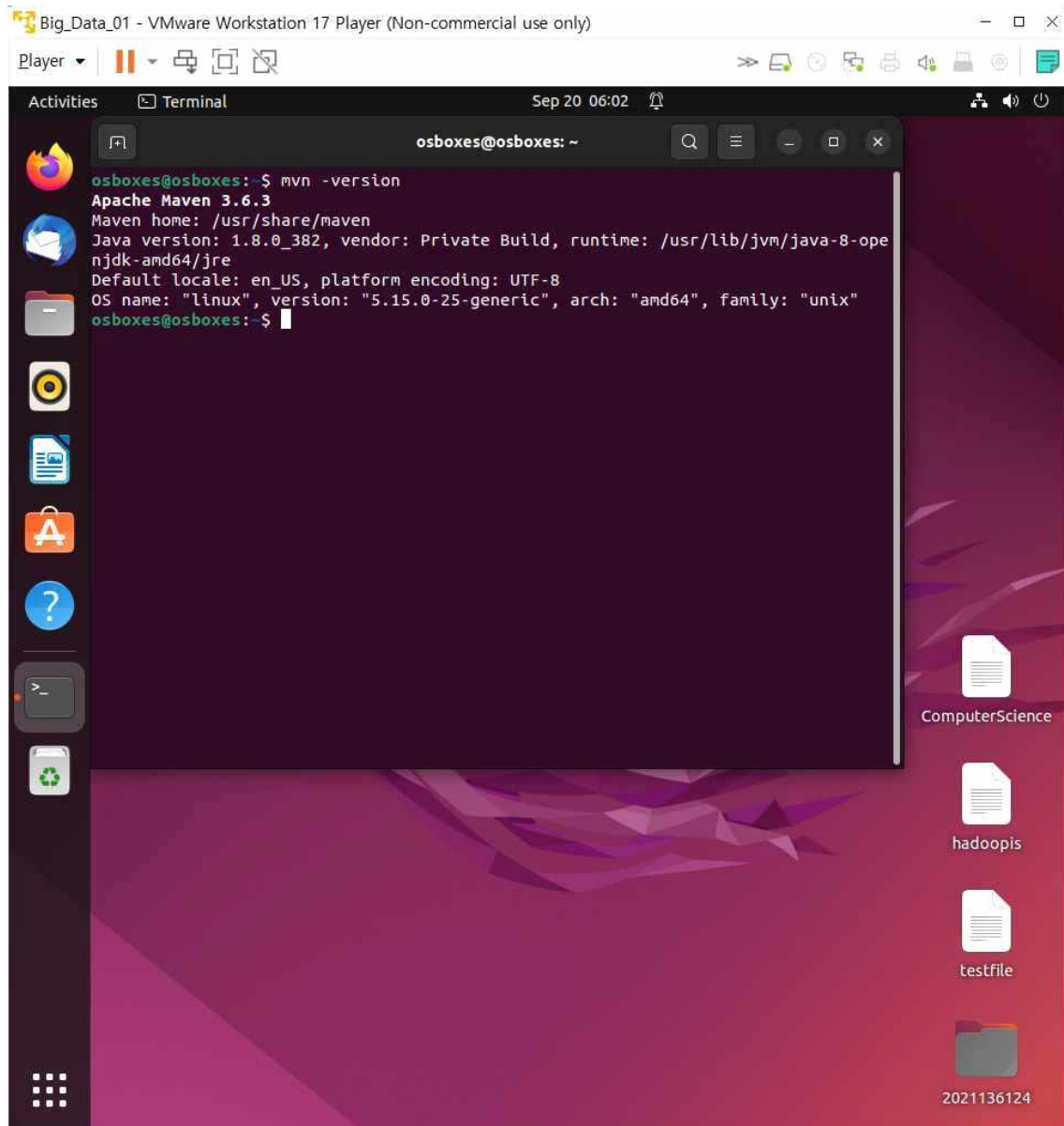
```
Big_Data_01 - VMware Workstation 17 Player (Non-commercial use only)
Player
osboxes@osboxes: ~
osboxes@osboxes:~$ hdfs dfs -put /home/Hadoop/Install/hadoop-3.3.1/etc/hadoop/core-site.xml /user/osboxes/test
osboxes@osboxes:~$ yarn jar /home/Hadoop/Install/hadoop-3.3.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar wordcount test output
2023-09-26 02:06:30,309 INFO client.DefaultNoHARMFaoverProxyProvider: Connecting to ResourceMa
nager at /0.0.0.0:8032
2023-09-26 02:06:31,168 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /
tmp/hadoop-yarn/staging/osboxes/.staging/job_1695706399451_0002
2023-09-26 02:06:31,808 INFO input.FileInputFormat: Total input files to process : 1
2023-09-26 02:06:32,041 INFO mapreduce.JobSubmitter: number of splits:1
2023-09-26 02:06:32,757 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_169570639945
1_0002
2023-09-26 02:06:32,757 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-09-26 02:06:33,511 INFO conf.Configuration: resource-types.xml not found
2023-09-26 02:06:33,512 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-09-26 02:06:34,684 INFO impl.YarnClientImpl: Submitted application application_169570639945
1_0002
2023-09-26 02:06:34,822 INFO mapreduce.Job: The url to track the job: http://osboxes:8088/proxy/
application_1695706399451_0002/
2023-09-26 02:06:34,824 INFO mapreduce.Job: Running job: job_1695706399451_0002
2023-09-26 02:06:56,694 INFO mapreduce.Job: Job job_1695706399451_0002 running in uber mode : fa
lse
2023-09-26 02:06:56,703 INFO mapreduce.Job: map 0% reduce 0%
2023-09-26 02:07:05,945 INFO mapreduce.Job: map 100% reduce 0%
2023-09-26 02:07:14,055 INFO mapreduce.Job: map 100% reduce 100%
2023-09-26 02:07:15,086 INFO mapreduce.Job: Job job_1695706399451_0002 completed successfully
2023-09-26 02:07:15,287 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=1236
FILE: Number of bytes written=547587
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=984
HDFS: Number of bytes written=894
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=6355
Total time spent by all reduces in occupied slots (ms)=5685
Total time spent by all map tasks (ms)=6355
Total time spent by all reduce tasks (ms)=5685
Total vcore-milliseconds taken by all map tasks=6355
```

이클립스 설치

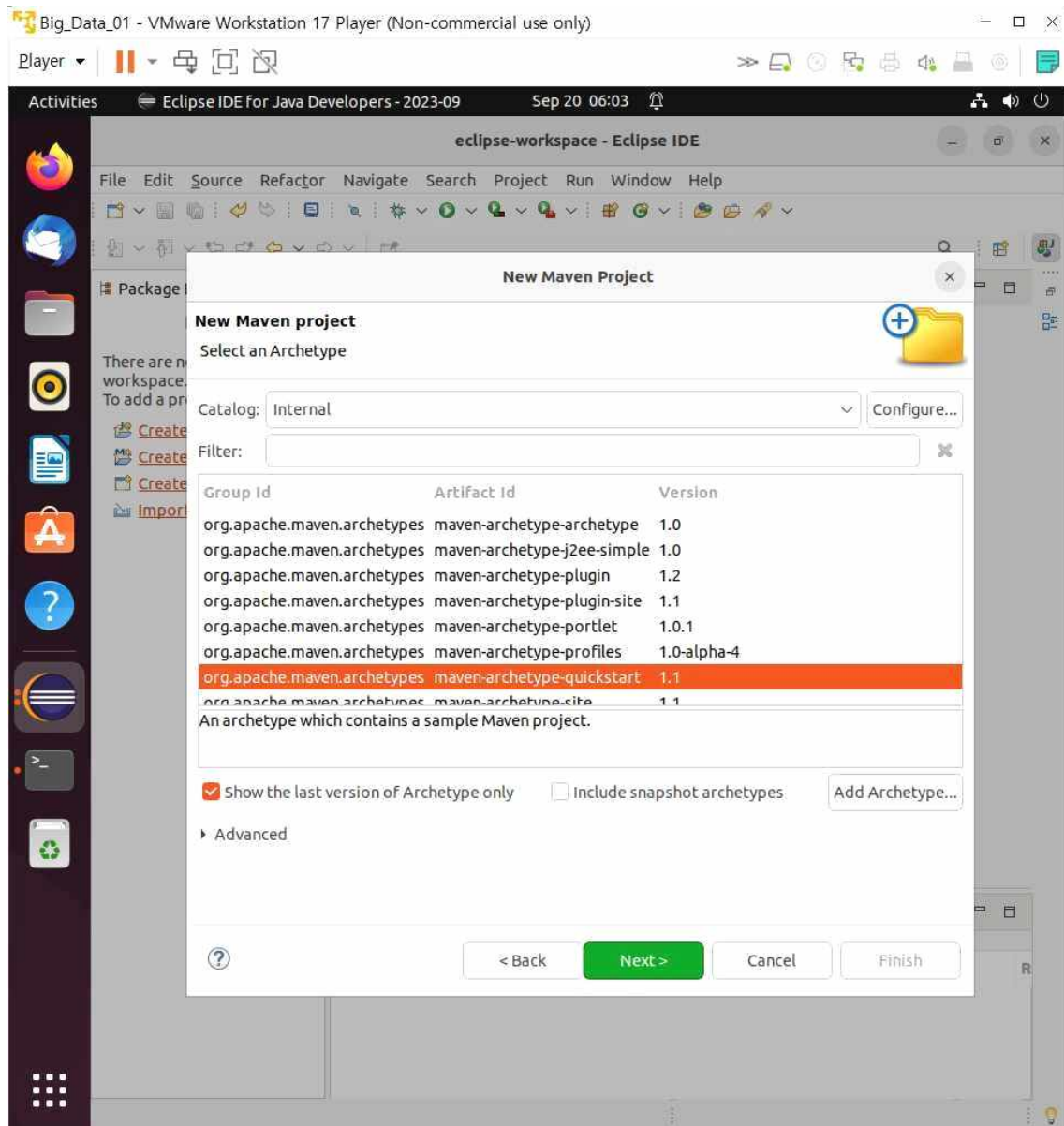


메이븐 확인

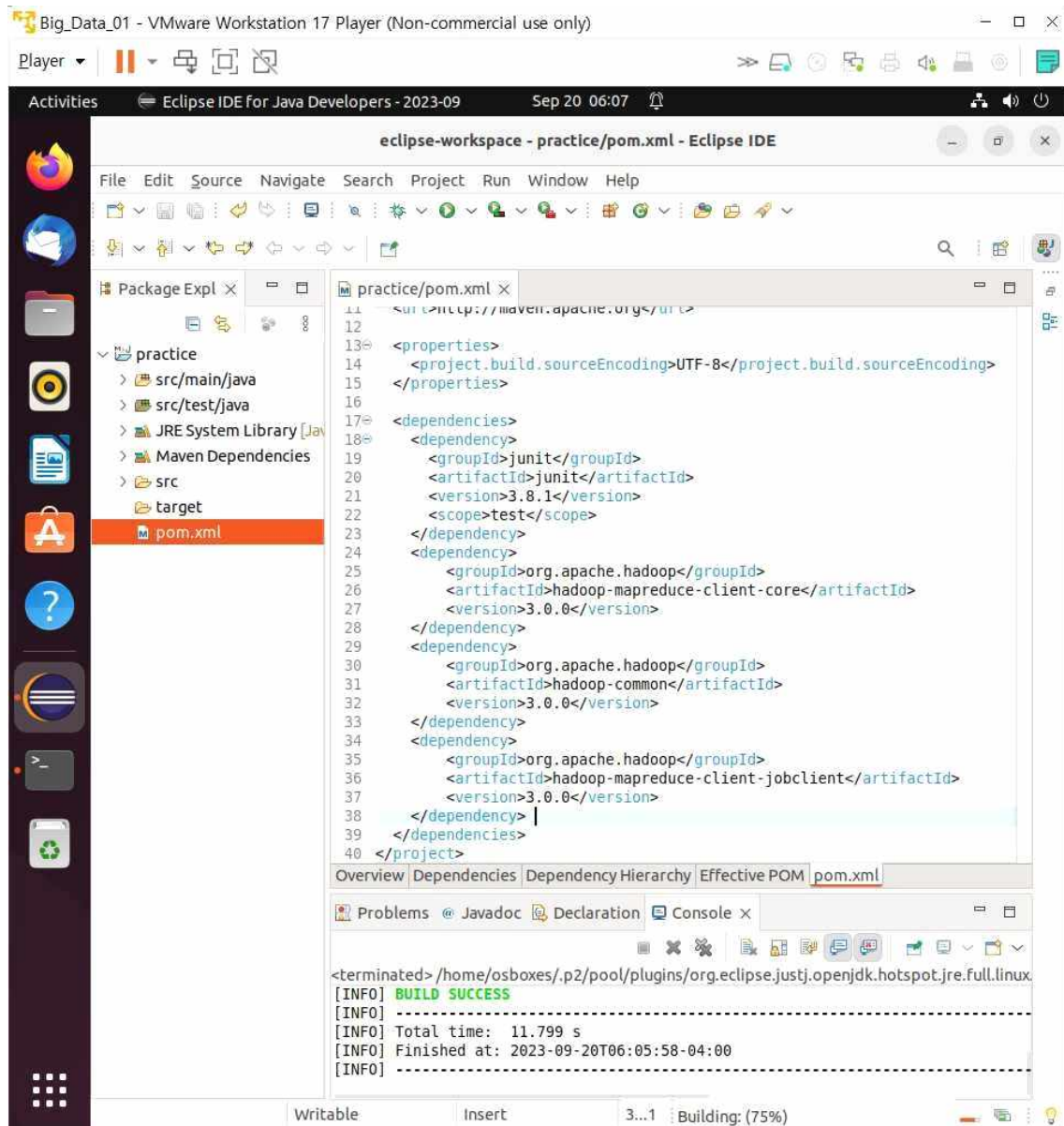




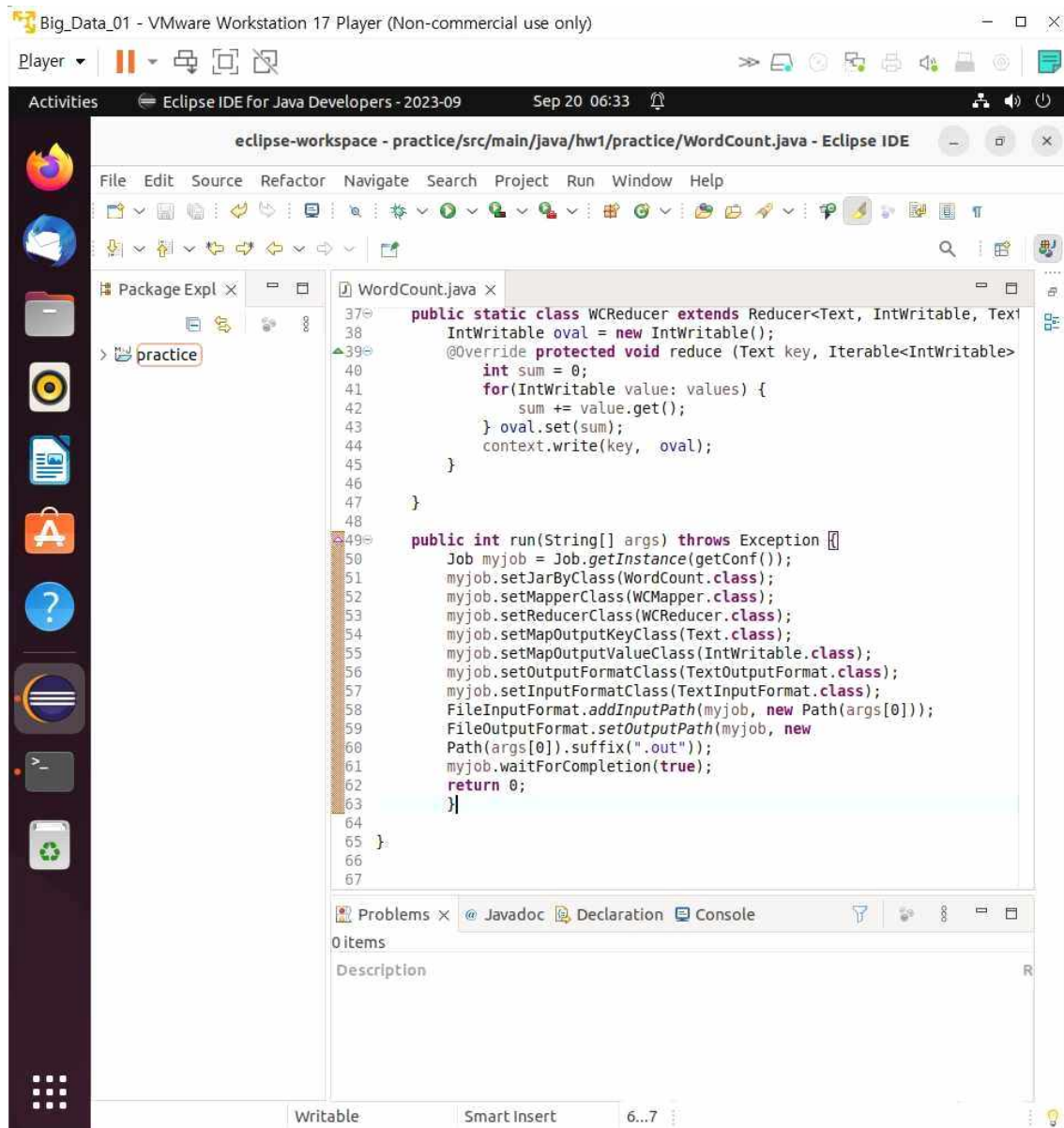
메이븐 프로젝트 생성



pom.xml 수정

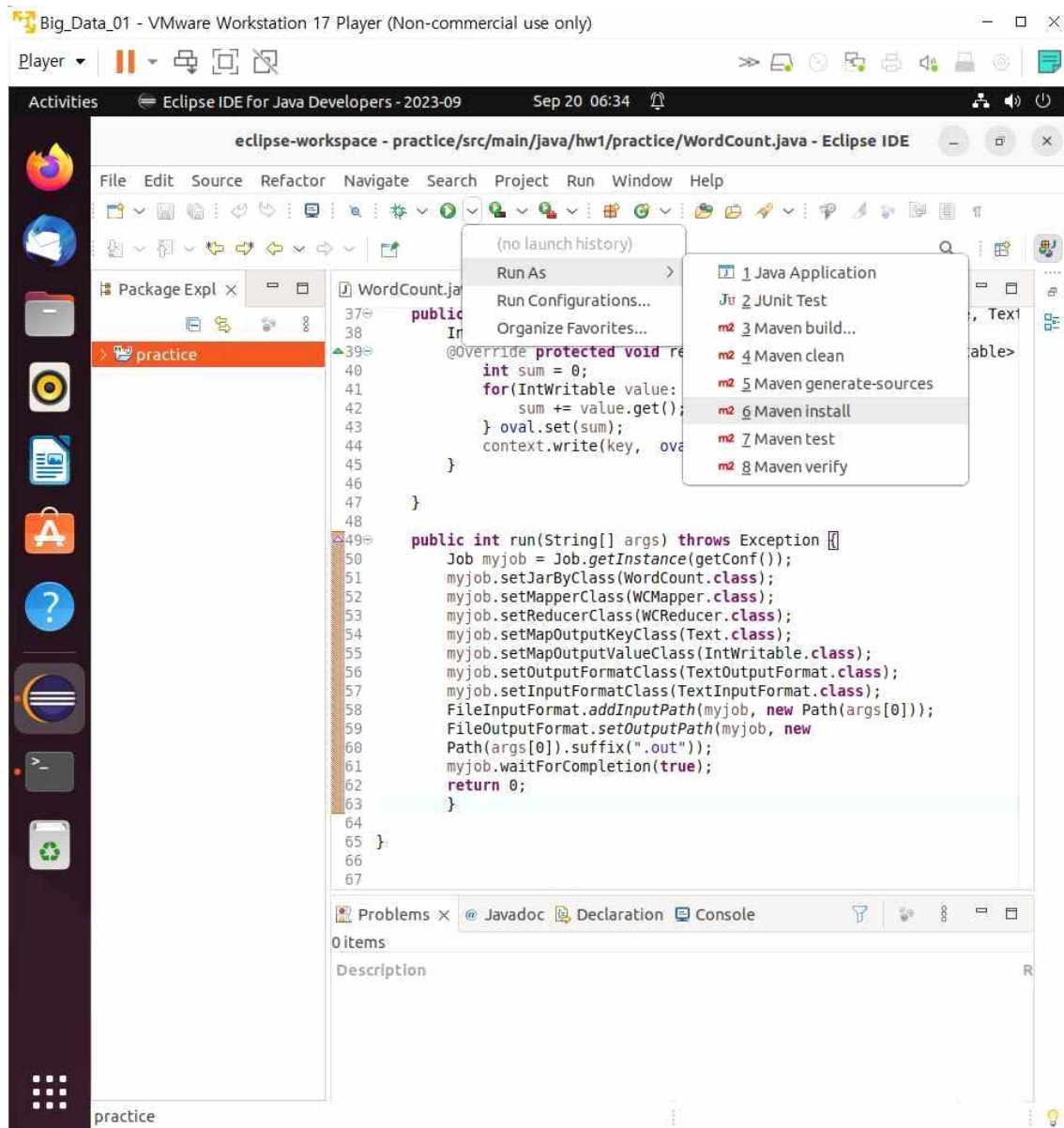


MapReduce 코드 작성

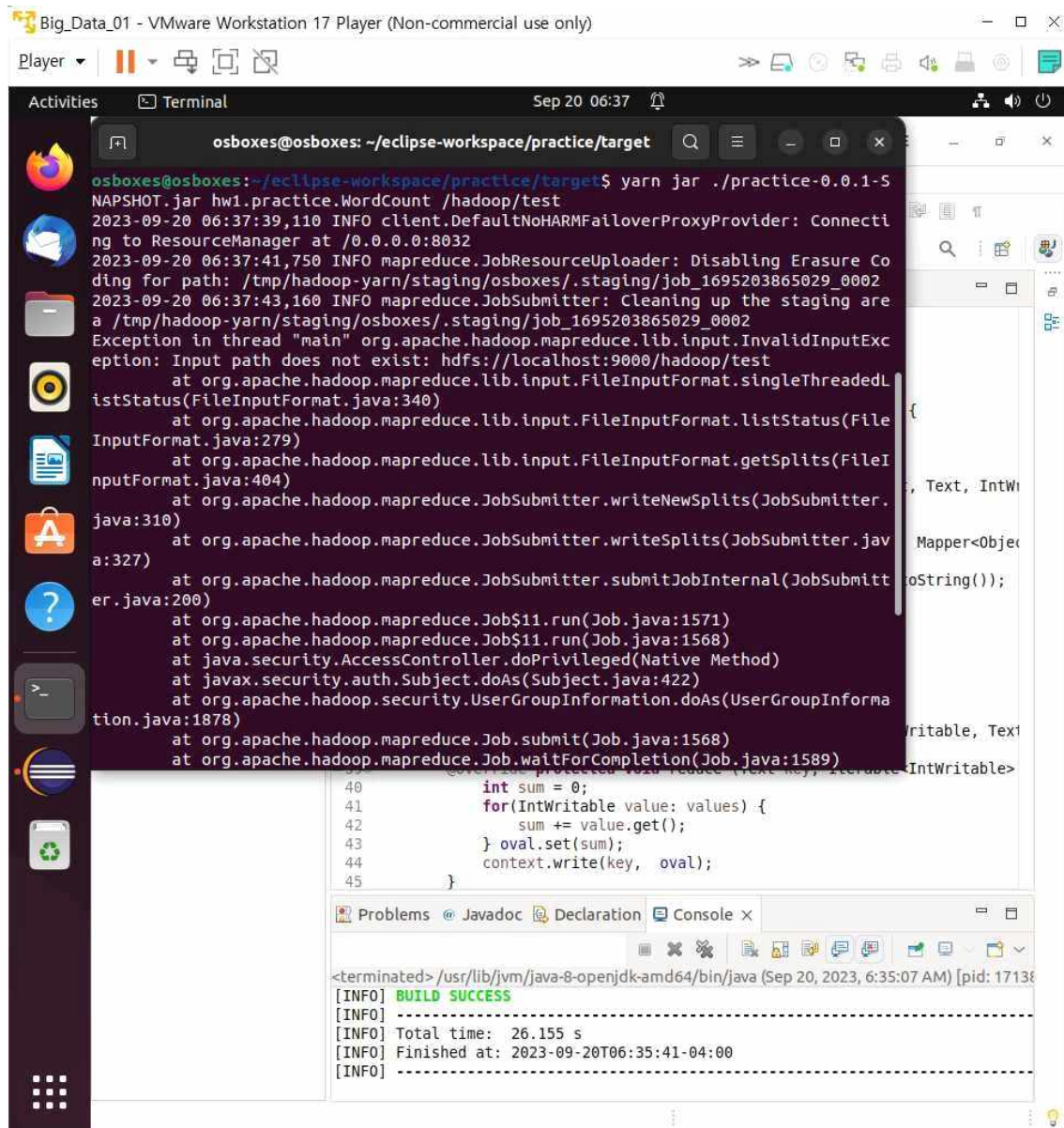


메이븐 컴파일

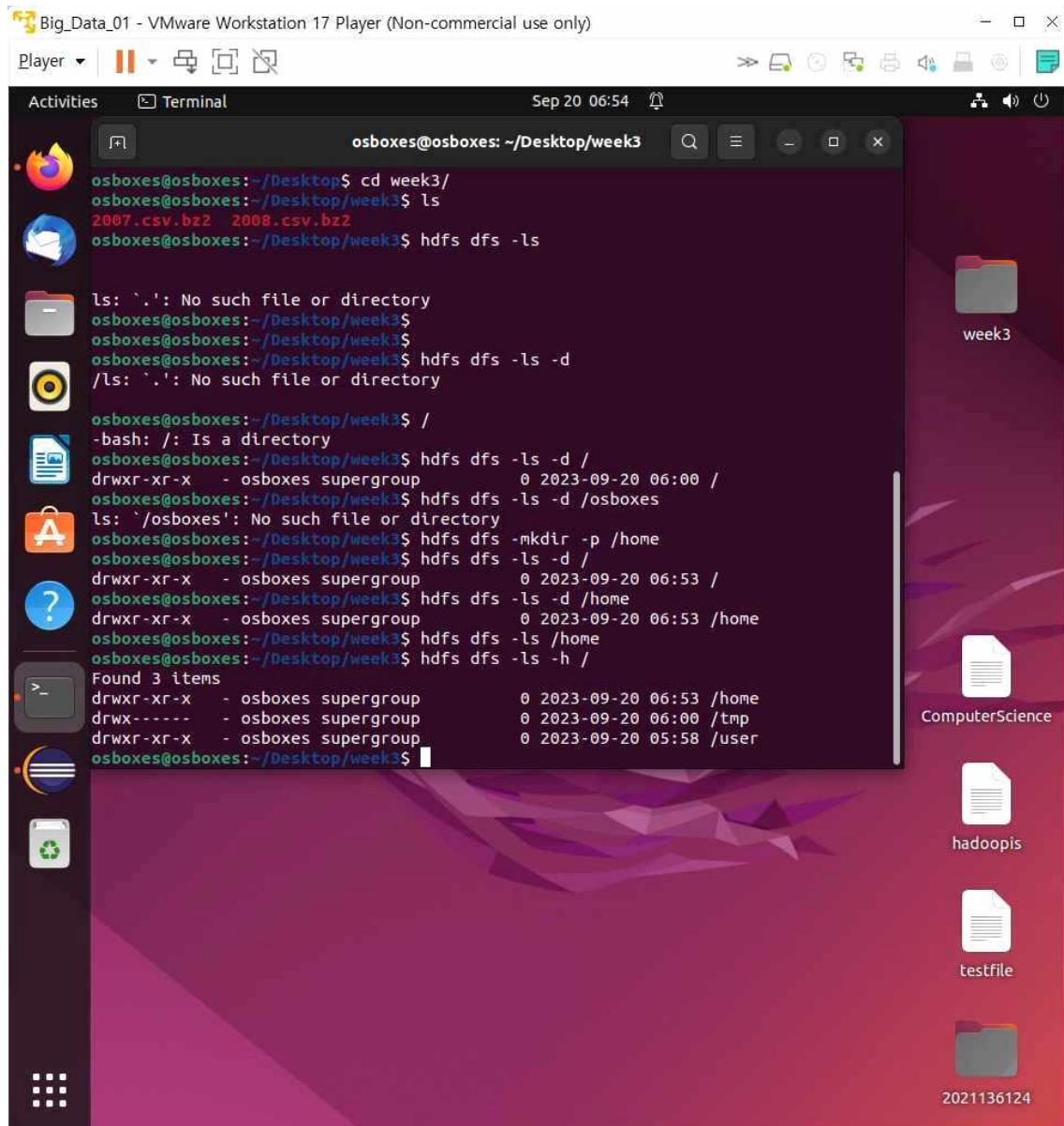




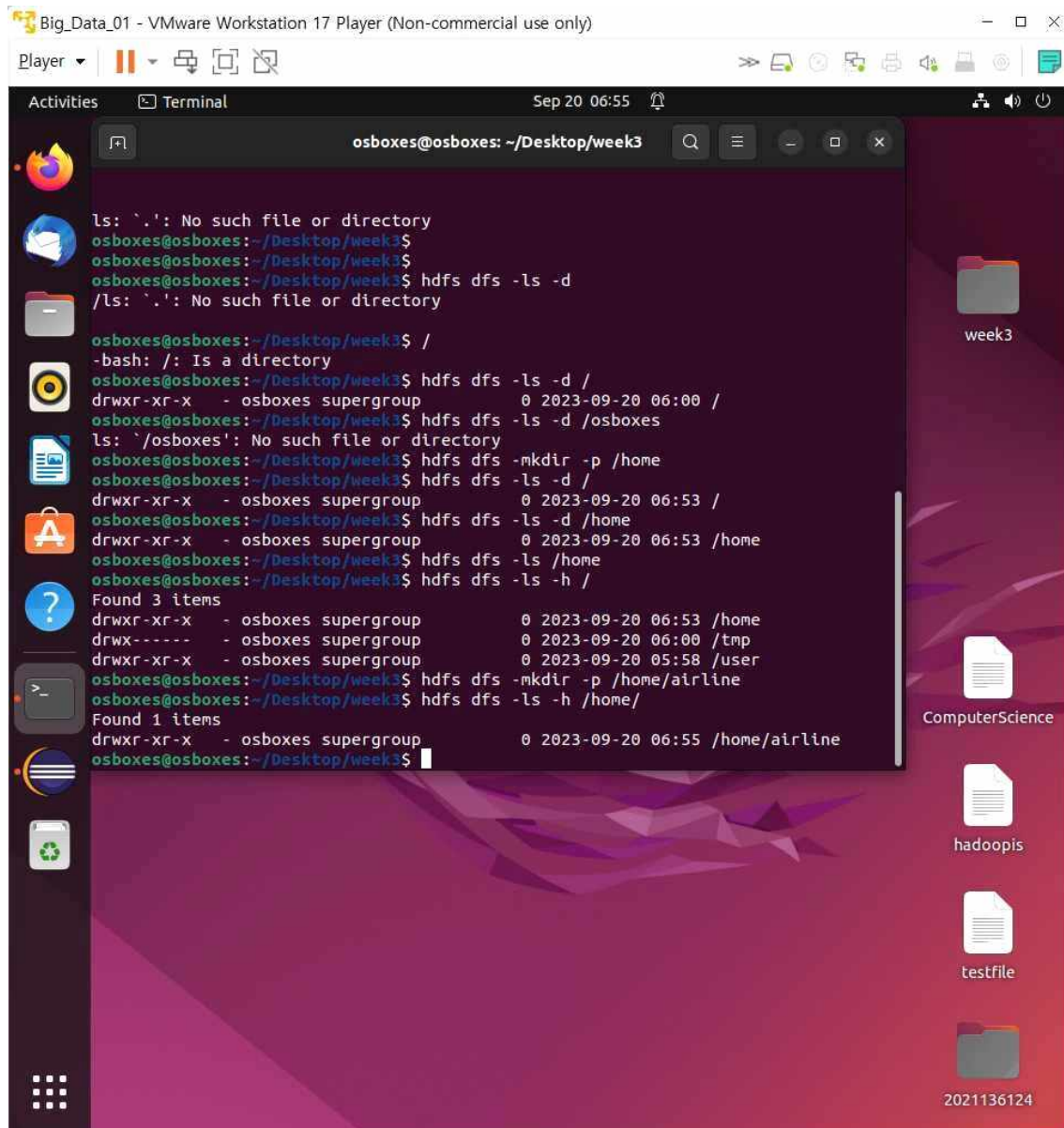
jar 파일 실행



사용자의 홈 디렉토리 생성

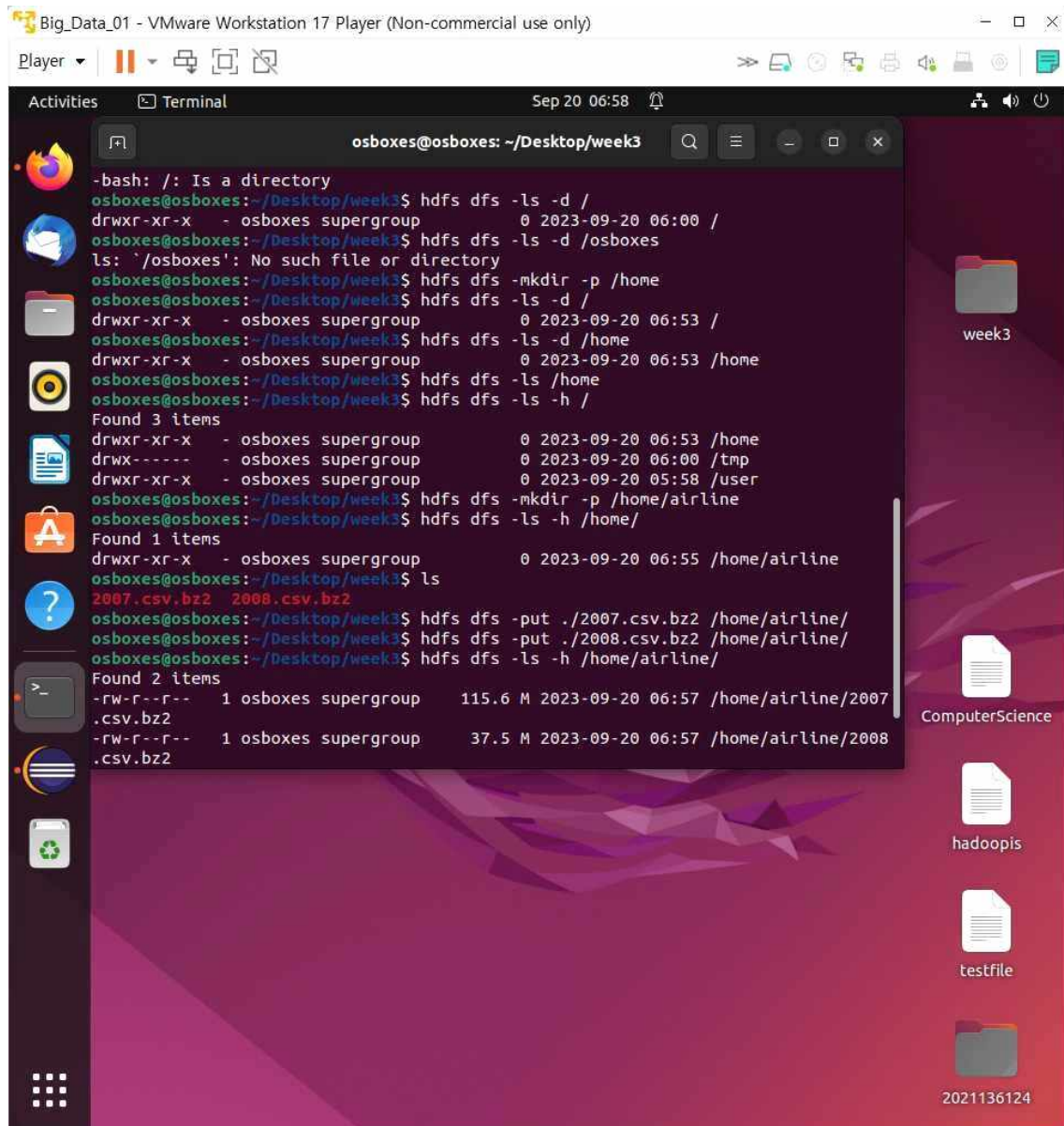


사용자의 홈 디렉토리에 airline 디렉토리 생성

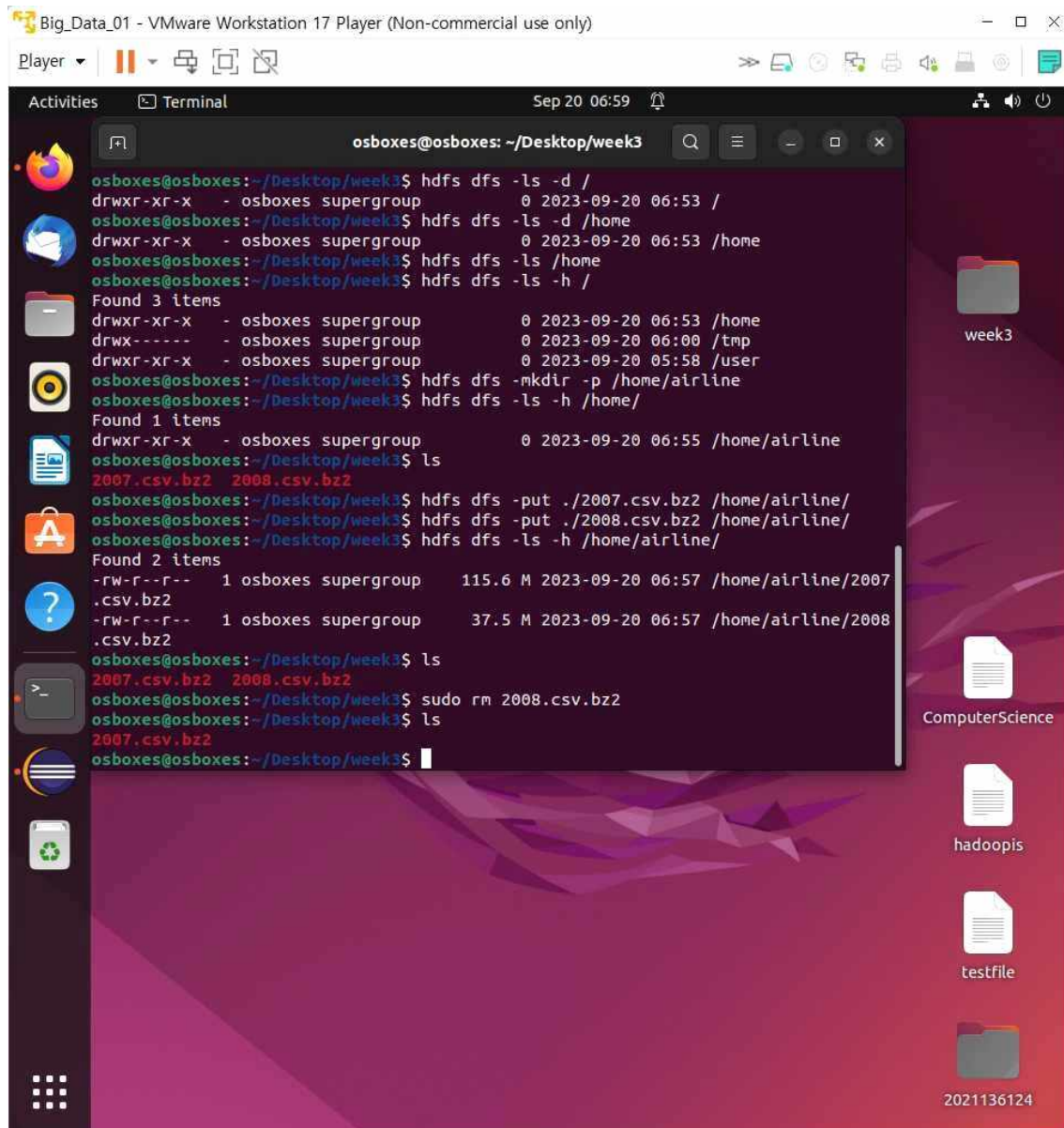


airline 디렉토리에 2008.csv 파일을 업로드 && airline 디렉토리에 2007.csv 파일을 업로드

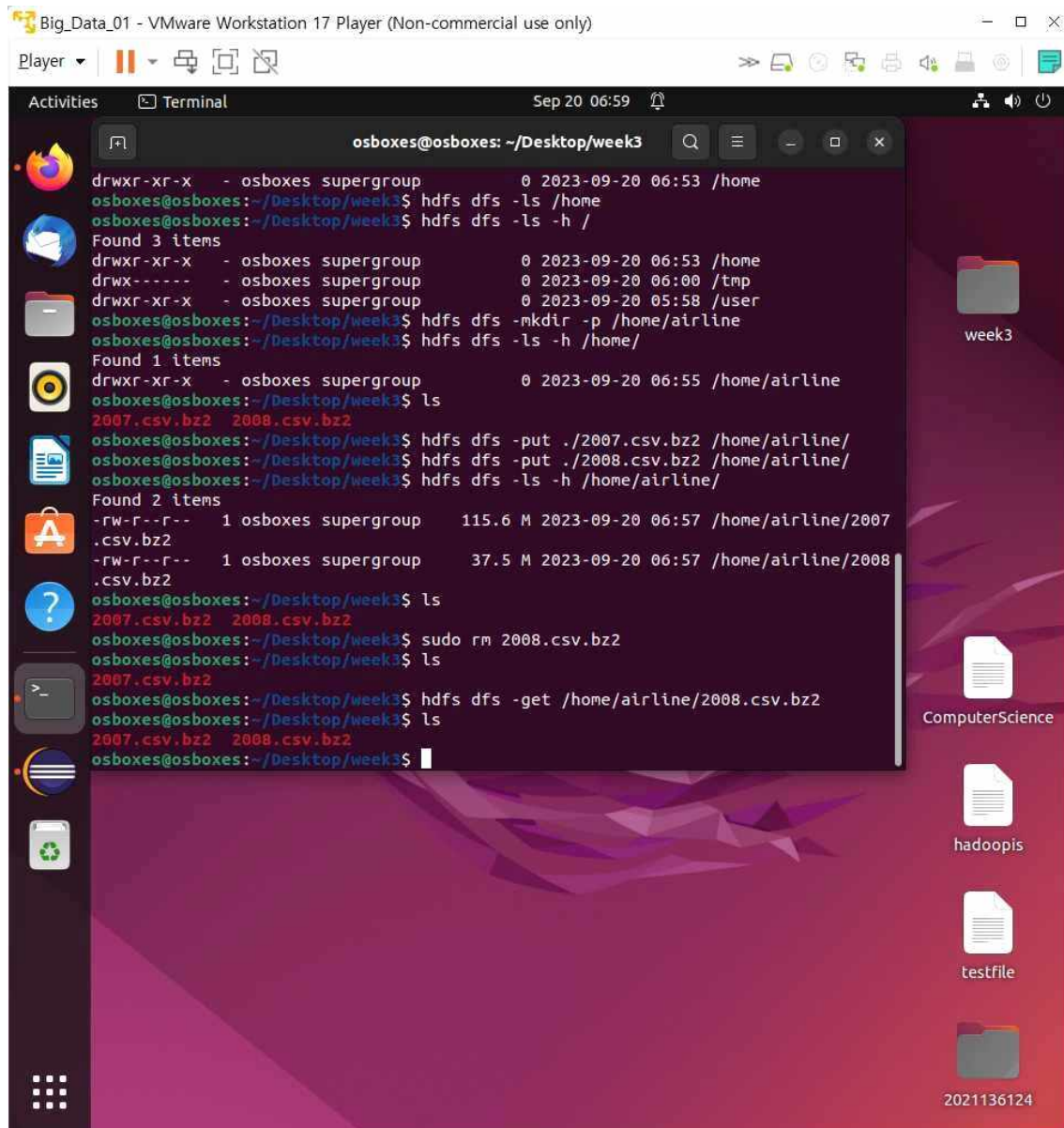




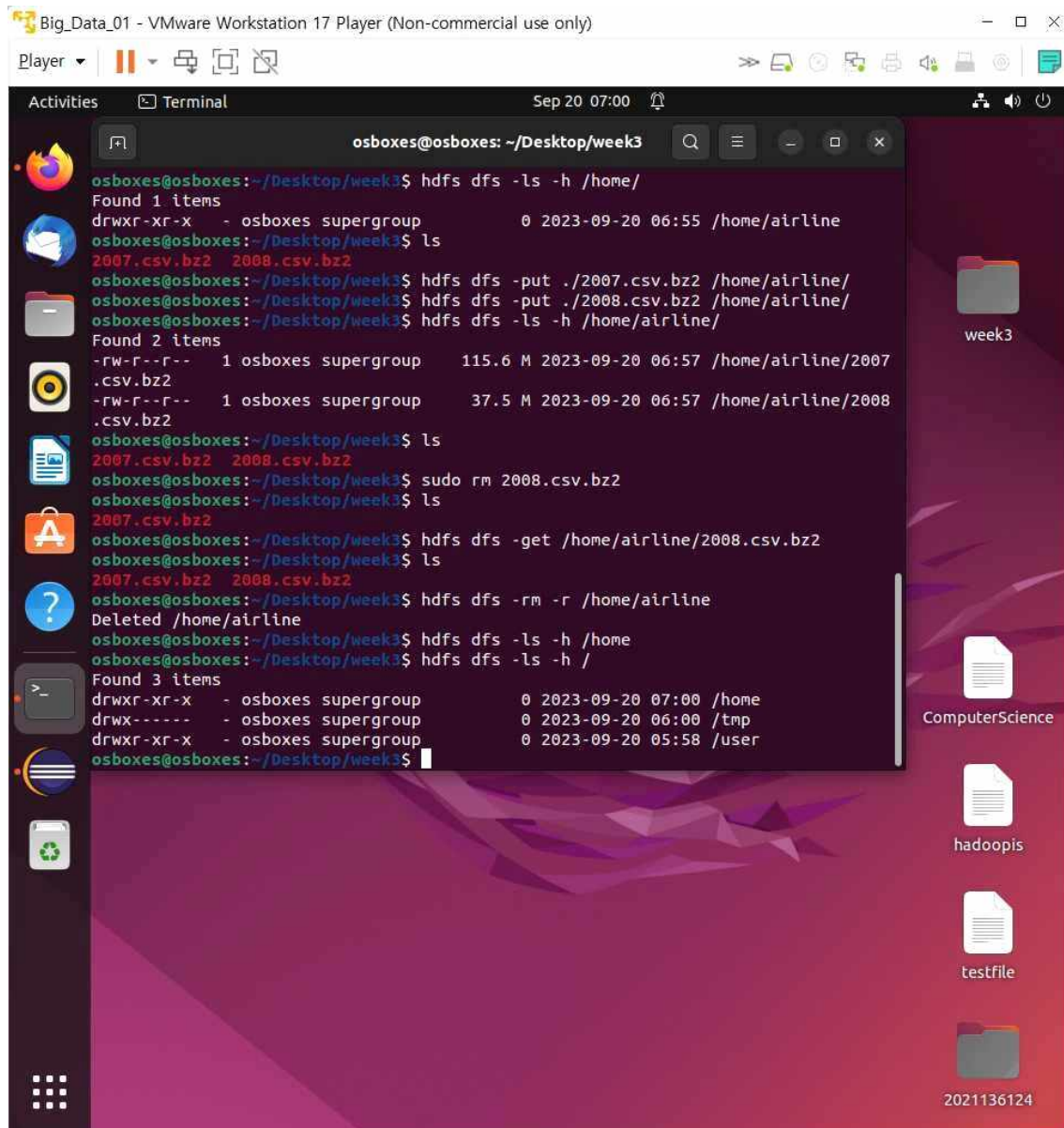
로컬의 2008.csv 파일 제거



HDFS의 2008.csv 파일을 로컬에 저장

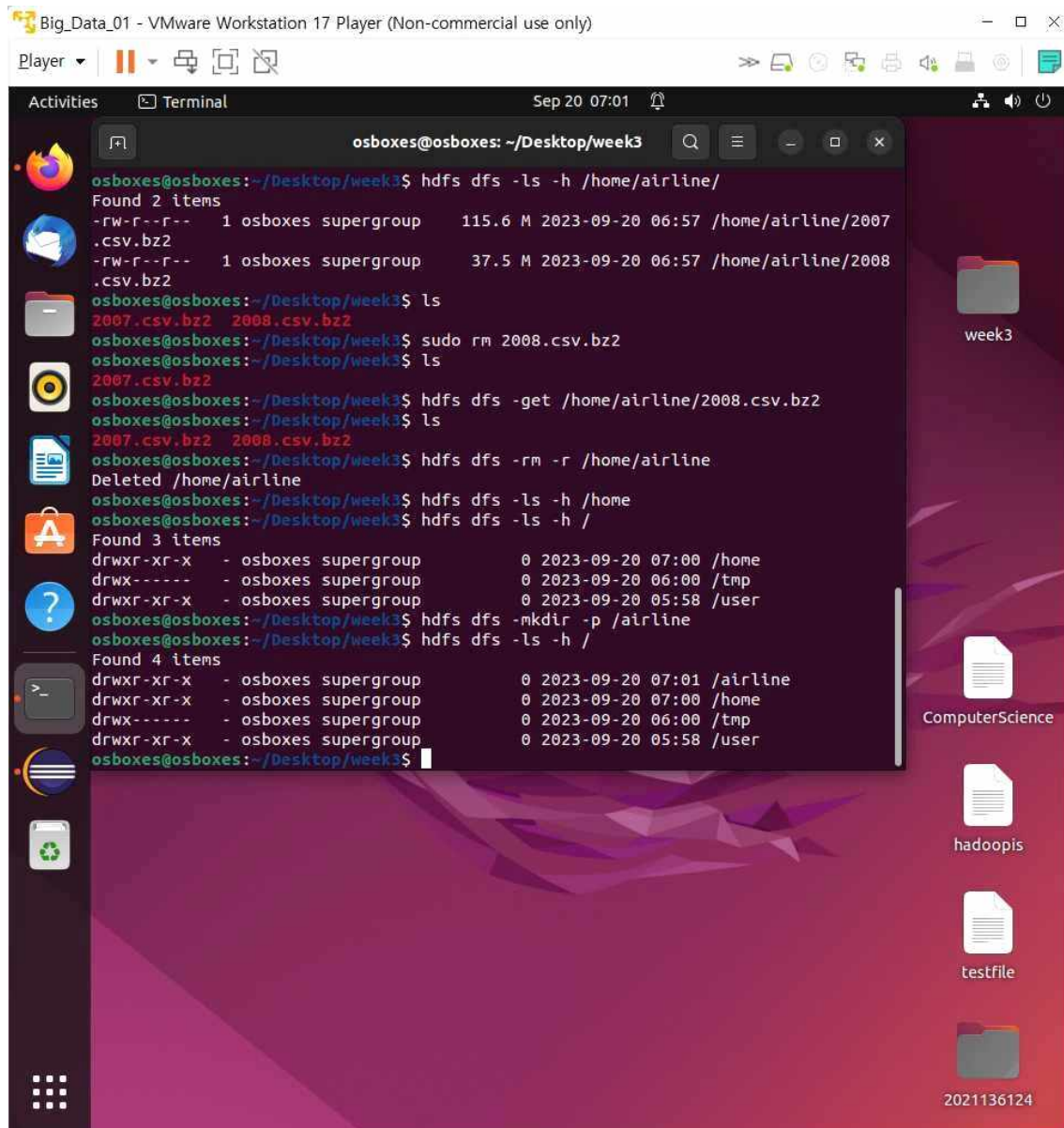


airline 디렉토리 삭제

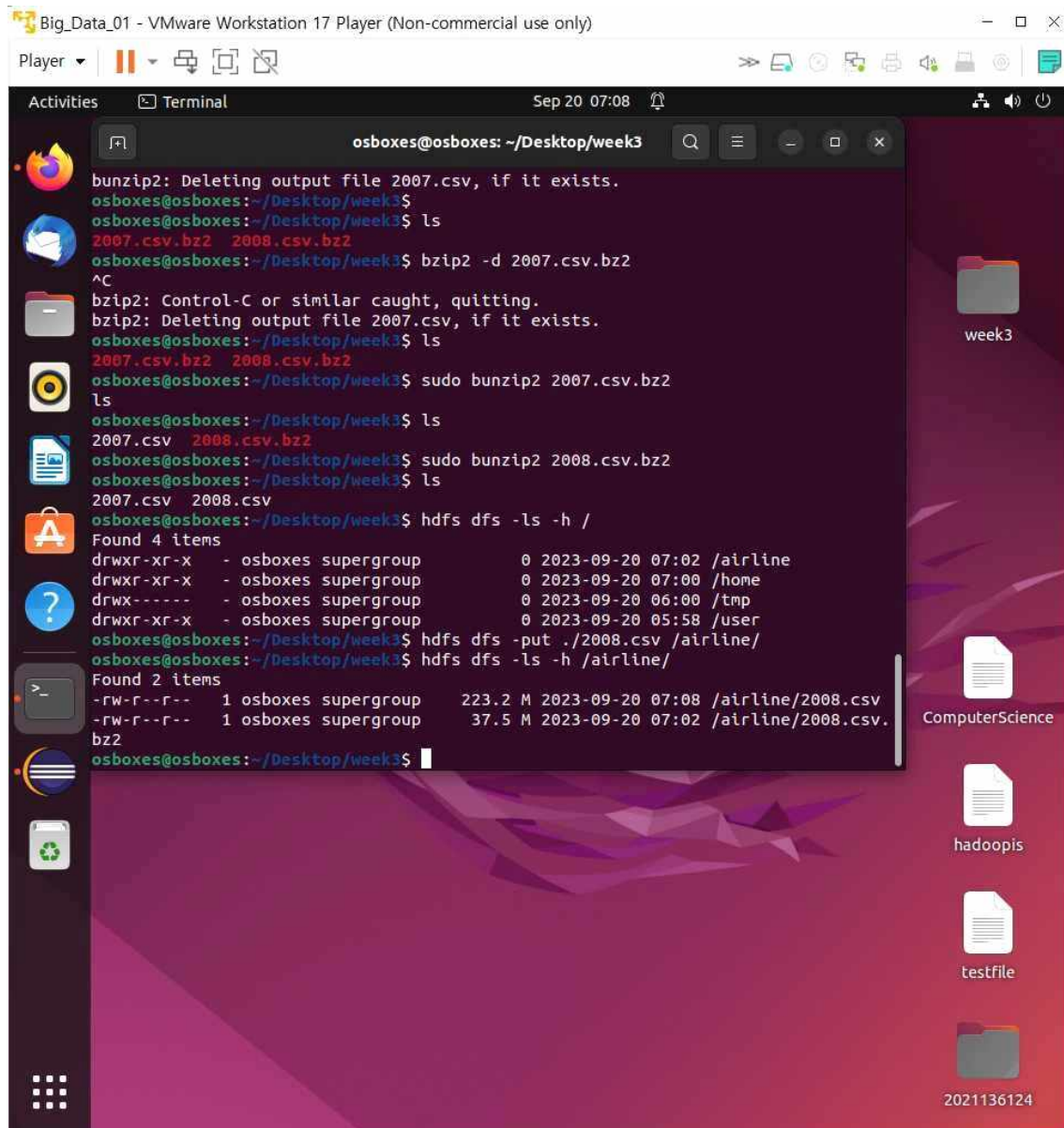


루트에 airline 디렉토리 생성

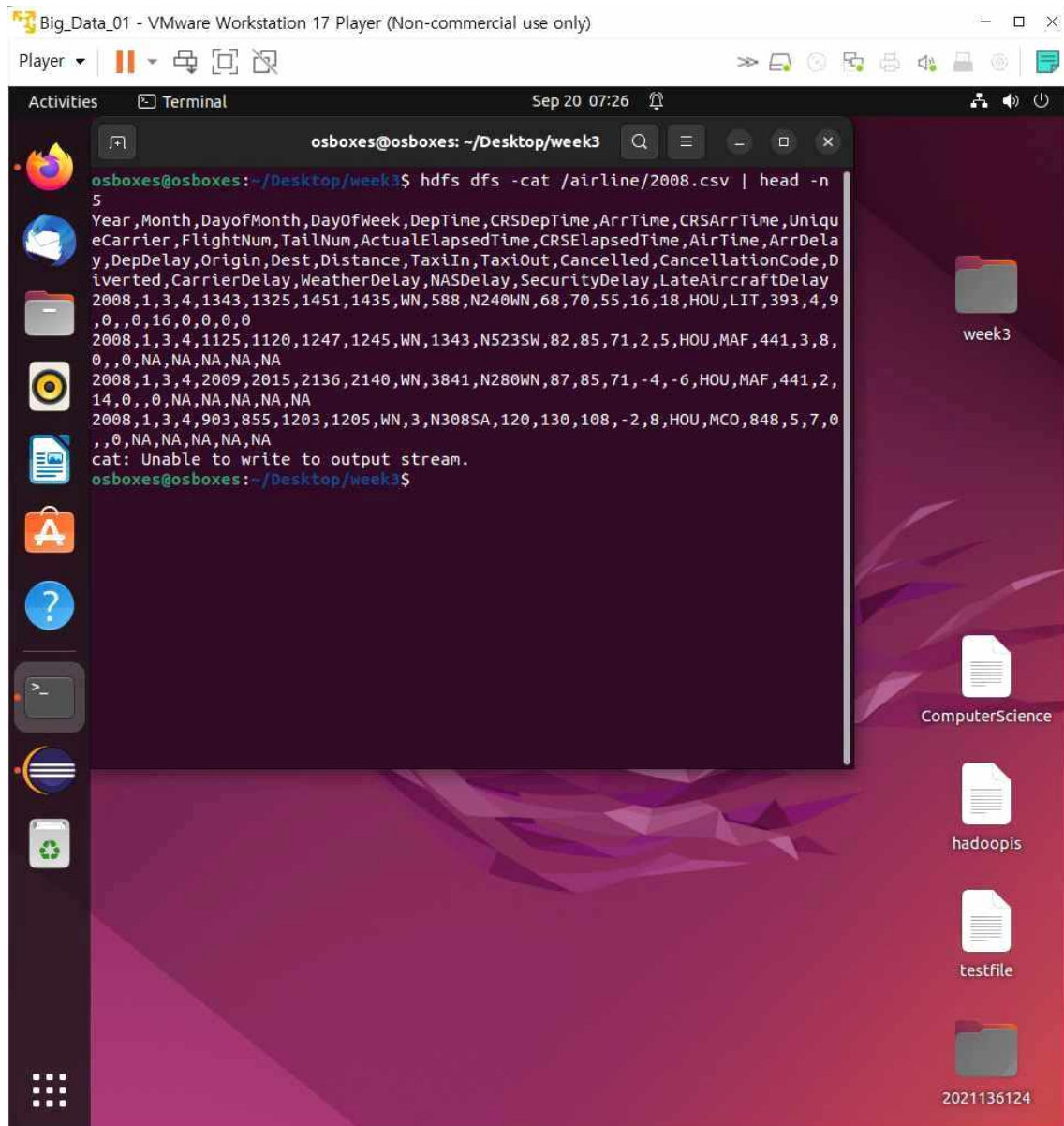




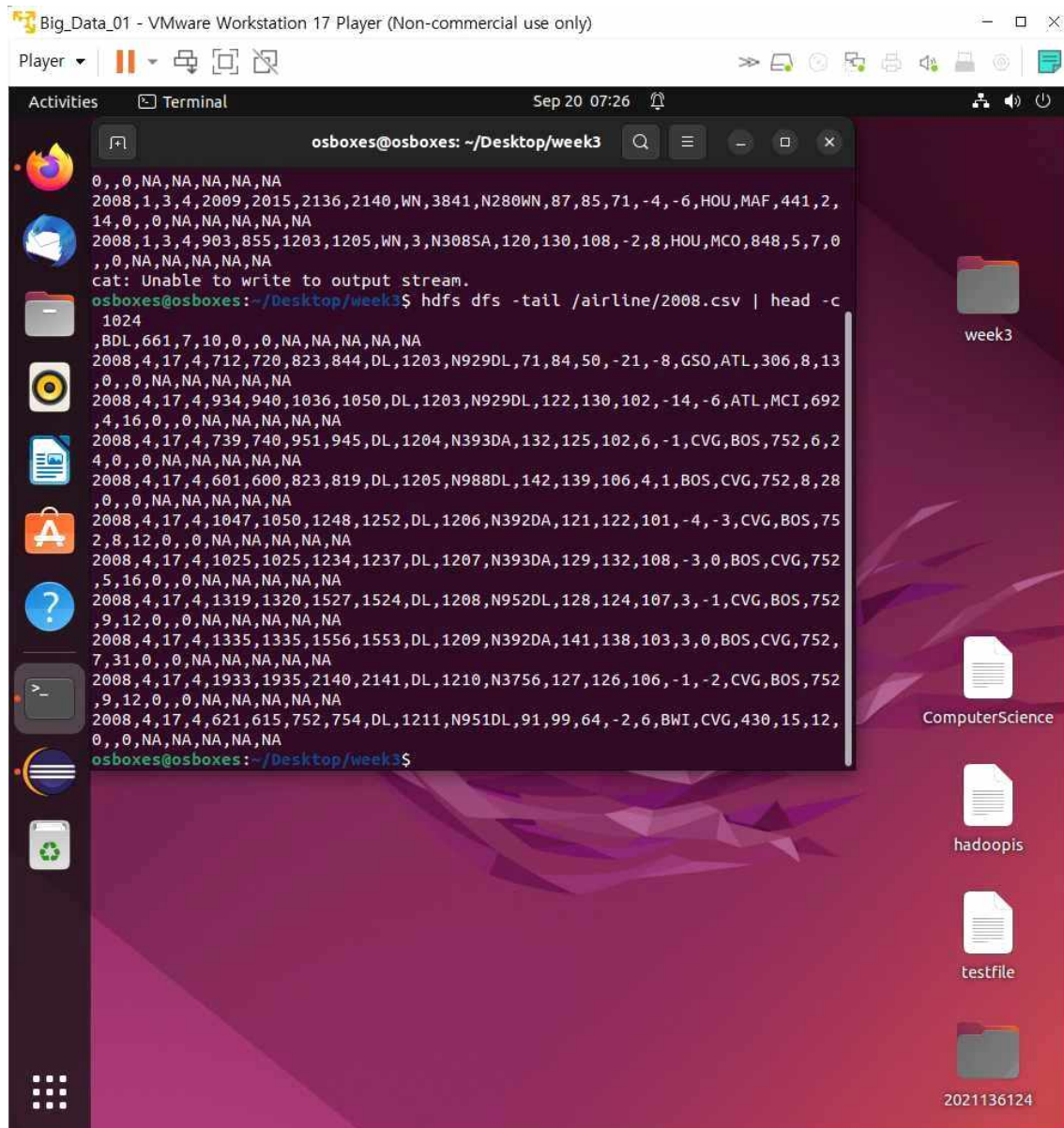
airline 디렉토리에 2008.csv 업로드



2008.csv 파일의 첫 5라인 출력

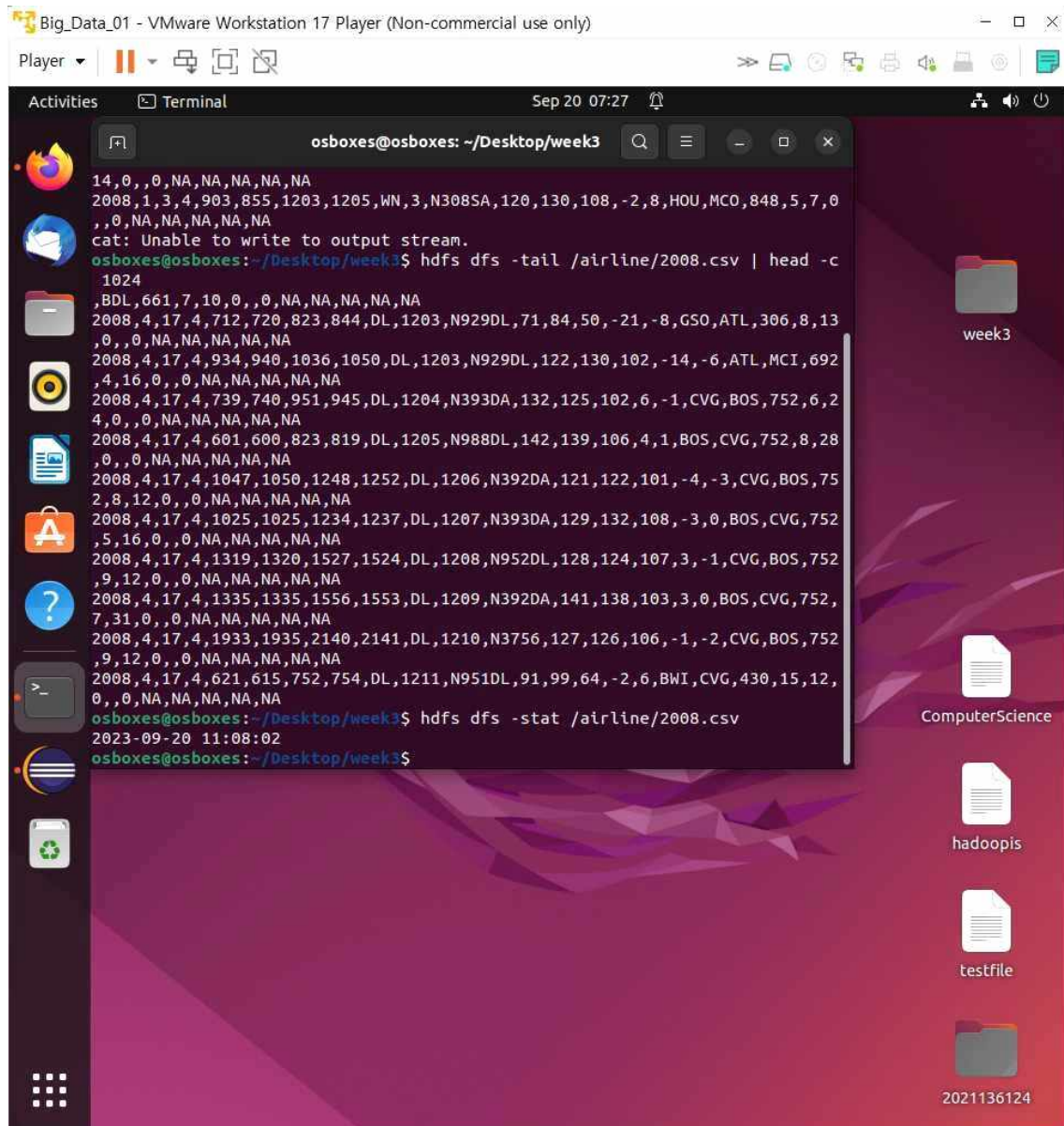


2008.csv 파일의 마지막 1KB를 출력



2008.csv 파일의 통계 정보 조회





2008.csv 파일의 복제 데이터 개수 변경 && 2008.csv 파일의 복제 수를 확인 &&  
2008.csv 파일의 복제 수를 1로 변경

