

5 주차 Homework

학번	32164420	이름	조정민
----	----------	----	-----

5 주차 수업 PPT 26 페이지의 Word Cont 예제.

※ 문제를 해결하여 제출하시오. 각 문제마다 실행 결과를 캡처하여 붙여넣는다.

1. HDFS 명령어를 통해 기존에 HDFS 에 생성된 폴더가 있는지 확인한다.

Answer:

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/bigdata$ hadoop fs -ls /
Found 3 items
drwxr-xr-x - jeongmin supergroup 0 2021-04-07 16:27 /input
drwxr-xr-x - jeongmin supergroup 0 2021-04-07 16:30 /output
drwxr-xr-x - jeongmin supergroup 0 2021-04-07 17:13 /output2
```

2. HDFS 에 기존 생성된 폴더와 파일을 모두 삭제한다.

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/bigdata$ hadoop fs -rm -r /input /output /output2
21/04/07 17:40:52 INFO fs.TrashPolicyDefault: Namenode trash configuration: Dele
tion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /input
21/04/07 17:40:52 INFO fs.TrashPolicyDefault: Namenode trash configuration: Dele
tion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output
21/04/07 17:40:52 INFO fs.TrashPolicyDefault: Namenode trash configuration: Dele
tion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output2
jeongmin@jeongmin:~/bigdata$ hadoop fs -ls /
jeongmin@jeongmin:~/bigdata$
```

3. HDFS 에 input 폴더를 생성한다.

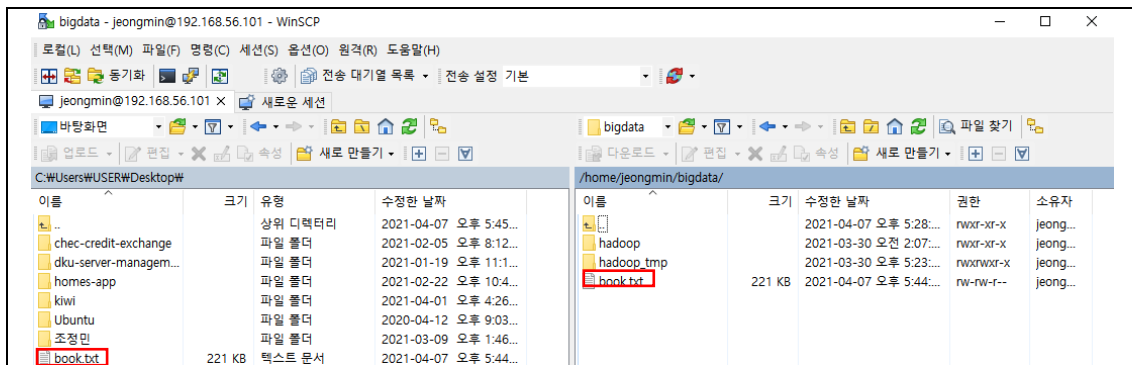
Answer:

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/bigdata$ hadoop fs -mkdir -p /input
jeongmin@jeongmin:~/bigdata$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - jeongmin supergroup 0 2021-04-07 17:42 /input
```

4. 파일은 구텐베르크 프로젝트(<https://www.gutenberg.org/>)에서 소설 하나를 다운로드 받아 book.txt 로 저장한 뒤, FTP 클라이언트를 이용해 네임서버로 이동시킨다.

Answer:

```
Font: consolas / Font size: 10
```



5. 네임서버로 이동시킨 book.txt 를 HDFS 에 생성한 input 폴더로 복사한다.

Answer:

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/bigdata$ hadoop fs -ls /input
Found 1 items
-rw-r--r--  3 jeongmin supergroup    225745 2021-04-07 17:51 /input/book.txt
```

6. WordCount 의 jar 파일을 실행한다. (본인이 직접 빌드하거나 샘플 파일을 이용)

Answer:

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/work$ hadoop jar hadoop-mapreduce-examples-2.5.2.jar wordcount /input /output
21/04/07 17:53:41 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
21/04/07 17:53:41 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
21/04/07 17:53:42 INFO input.FileInputFormat: Total input paths to process : 1
21/04/07 17:53:42 INFO mapreduce.JobSubmitter: number of splits:1
```

7. 실행 결과를 확인한다.

Answer:

```
Font: consolas / Font size: 10
jeongmin@jeongmin:~/work$ hadoop fs -ls /output
Found 2 items
-rw-r--r--  3 jeongmin supergroup      0 2021-04-07 17:53 /output/_SUCCESS
-rw-r--r--  3 jeongmin supergroup 81305 2021-04-07 17:53 /output/part-r-00000000
jeongmin@jeongmin:~/work$ hadoop fs -cat /output/part-r-00000000
"Defects,"      1
"Information    1
"Plain 2
"Project        5
"Right 1
#65012] 1
$5,000) 1
&c., 3
'AS-IS',      1
("the 1
($1 1
(801) 1
(God 1
(Mr. 1
```