5 주차 Homework			
학번	32151671	이름	박민혁

5 주차 수업 PPT 26 페이지의 Word Cont 예제.

- ※ 문제를 해결하여 제출하시오. 각 문제마다 실행 결과를 캡쳐하여 붙여넣는다.
- 1. HDFS 명령어를 통해 기존에 HDFS 에 생성된 폴더가 있는지 확인한다.

Answer:

```
hadoop@hadoop-name:~/Desktop$ hadoop fs -ls /
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2021-04-07 22:08 /input
drwxr-xr-x - hadoop supergroup 0 2021-04-07 22:09 /output
drwxr-xr-x - hadoop supergroup 0 2021-04-07 22:09 /output2
```

2. HDFS 에 기존 생성된 폴더와 파일을 모두 삭제한다.

```
hadoop@hadoop-name:~/Desktop$ hadoop fs -rm -r /input /output /output2
21/04/07 22:11:00 INFO fs.TrashPolicyDefault: Namenode trash configuration: Del
etion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /input
21/04/07 22:11:00 INFO fs.TrashPolicyDefault: Namenode trash configuration: Del
etion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output
21/04/07 22:11:00 INFO fs.TrashPolicyDefault: Namenode trash configuration: Del
etion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output2
```

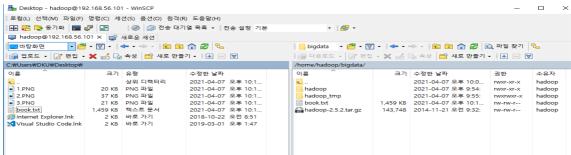
3. HDFS 에 input 폴더를 생성한다.

Answer:

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -mkdir -p /input
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2021-04-07 22:12 /input
hadoop@hadoop-name:~/bigdata$
```

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

Answer:



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

Answer:

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -put book.txt /input
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /input
Found 1 items
-rw-r--r-- 3 hadoop supergroup 1493884 2021-04-07 22:27 /input/book.txt
```

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

```
hadoop@hadoop-name:~/bigdata/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-m
apreduce-examples-2.5.2.jar wordcount /input /output
21/04/07 22:31:09 INFO Configuration.deprecation: session.id is deprecated. Ins
tead, use dfs.metrics.session-id
21/04/07 22:31:09 INFO jvm.JvmMetrics: Initializing JVM Metrics with processNam
e=JobTracker, sessionId=
21/04/07 22:31:09 INFO input.FileInputFormat: Total input paths to process : 1
21/04/07 22:31:10 INFO mapreduce.JobSubmitter: number of splits:1
21/04/07 22:31:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_l
ocal1592369626 0001
```

7. 실행 결과를 확인한다.

Answer¹

```
hadoop@hadoop-name:~/bigdataS hadoop fs -ls /output
Found 2 items
-rw-r--r-- 3 hadoop supergroup
                                     0 2021-04-07 22:31 /output/_SUCCESS
                                300073 2021-04-07 22:31 /output/part-r-000
- rw-r--r--
         3 hadoop supergroup
00
hadoop@hadoop-name:~/bigdata$ hadoop fs -cat /output/part-r-00000
invasion
                 6
invasion,
                  3
invasion.
                 3
inventory,
                  1
inventory.
investigate
investigate.
                  1
investigated
                  2
investigating
                  1
investigation
                  2
investigation,
                  3
investigations
                  4
investment
                  1
invitation?of
invited 2
invoiced
                  1
involve 1
involved
                  36
involved,
involving
                 10
iron
ironclad
ironically,
irregular
                  2
irreverently
is
         247
is372
island 32cat: Filesystem closed
```