

5 주차 Homework

학번	32170699	이름	김수연
----	----------	----	-----

5 주차 수업 PPT 26 페이지의 Word Count 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

```
hadoop@hadoop-name:~/work$ hadoop fs -ls /
Found 4 items
drwxr-xr-x - hadoop supergroup 0 2021-04-11 22:07 /bigdata
drwxr-xr-x - hadoop supergroup 0 2021-04-11 22:17 /input
drwxr-xr-x - hadoop supergroup 0 2021-04-11 22:20 /output
drwxr-xr-x - hadoop supergroup 0 2021-04-11 23:40 /output2
```

2. HDFS 에 기존 생성된 폴더와 파일을 모두 삭제한다.

Font: consolas / Font size: 10

```
hadoop@hadoop-name:~/work$ hadoop fs -rm -r /input /output /output2
21/04/12 00:04:21 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /input
21/04/12 00:04:21 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output
21/04/12 00:04:21 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /output2
```

3. HDFS 에 input 폴더를 생성한다.

Answer:

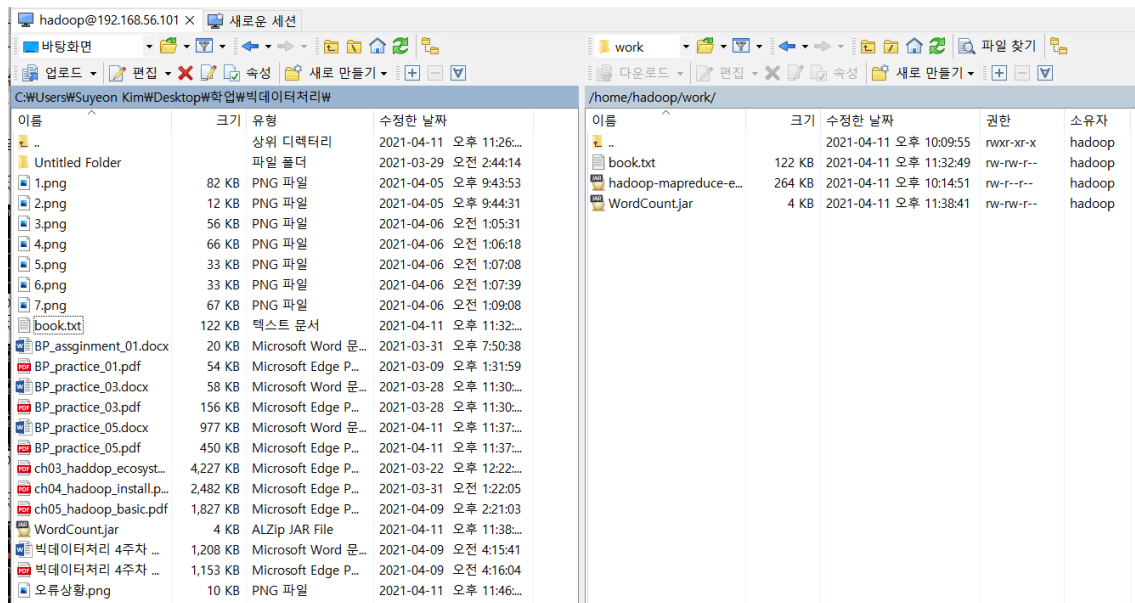
Font: consolas / Font size: 10

```
hadoop@hadoop-name:~/work$ hadoop fs -mkdir /input
hadoop@hadoop-name:~/work$ ls
hadoop-mapreduce-examples-2.5.2.jar WordCount.jar
hadoop@hadoop-name:~/work$ hadoop fs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup 0 2021-04-11 22:07 /bigdata
drwxr-xr-x - hadoop supergroup 0 2021-04-12 00:05 /input
hadoop@hadoop-name:~/work$
```

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

```
hadoop@hadoop-name:~/work$ hadoop fs -put book.txt /input
hadoop@hadoop-name:~/work$ hadoop fs -ls /input
Found 1 items
-rw-r--r--  3 hadoop supergroup      124670 2021-04-12 00:08 /input/book.txt
```

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

Answer:

Font: consolas / Font size: 10

```
hadoop@hadoop-name:~/work$ hadoop jar WordCount.jar WordCount /input /output2
21/04/11 23:40:54 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
21/04/11 23:40:54 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
21/04/11 23:40:54 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
21/04/11 23:40:55 INFO input.FileInputFormat: Total input paths to process : 1
21/04/11 23:40:55 INFO mapreduce.JobSubmitter: number of splits:1
21/04/11 23:40:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1378488524_0001
21/04/11 23:40:55 WARN conf.Configuration: file:/home/hadoop/bigdata/hadoop_tmp/tmp/mapred/staging/hadoop_p1378488524/.staging/job_local1378488524_0001/job.xml:an attempt to override final parameter: mapreduce.job.end-notification.max.retry.interval; Ignoring.
21/04/11 23:40:55 WARN conf.Configuration: file:/home/hadoop/bigdata/hadoop_tmp/tmp/mapred/staging/hadoop_p1378488524/.staging/job_local1378488524_0001/job.xml:an attempt to override final parameter: mapreduce.job.end-notification.max.attempts; Ignoring.
```

7. 실행 결과를 확인한다.

Answer:

Font: consolas / Font size: 10

```
hadoop@hadoop-name:~/work$ hadoop fs -cat /output2/part-r-00000
(BIS), 1
(ECCN) 1
(TSU) 1
(see 1
5D002.C.1, 1
740.13) 1
<http://www.wassenaar.org/> 1
Administration 1
Apache 1
BEFORE 1
BIS 1
Bureau 1
Commerce, 1
Commodity 1
Control 1
Core 1
Department 1
ENC 1
Exception 1
Export 2
For 1
Foundation 1
Government 1
Hadoop 1
Hadoop, 1
Industry 1
Jetty 1
License 1
Number 1
Regulations, 1
```