빅데이터 분석시스템 개발

빅데이터 분석모듈 개발하기

양용민

hadoop-streaming 라이브러리 복사

```
[root@hadoop101 ~]# cd ~
[root@hadoop101 ~]# mkdir MapReduce
[root@hadoop101 ~]# cp $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-2.10.1.jar  /root/MapReduce
```

mapper, reducer 작성

word_mapper.py

```
[root@hadoop101 ~]# cd /root/MapReduce
[root@hadoop101 MapReduce]# vi word_mapper.py
```

```
[root@hadoop101 MapReduce]# vi word_mapper.py
     1 #!/usr/bin/python
     2 import sys
     3
     4 for line in sys.stdin:
     5         line = line.strip()
     6         words = line.split()
     7
     8         for word in words:
     9                 print('{}\t{}'.format(word, 1))
```

```
~
:set fileformat=unix
```

word_reducer.py

```
[root@hadoop101 MapReduce]# vi word_reducer.py
     1 #!/usr/bin/python
     2 import sys
     3 word, count = None, 0
     4
     5 for line in sys.stdin:
     6         fields = line.strip().split('\t')
     7
     8         if fields[0] != word:
     9                 if word is not None:
    10                         print('{}\t{}'.format(word, count))
    11
    12                 word, count = fields[0], 0
    13
    14         count += 1
    15
    16 print('{}\t{}'.format(word, count))
```

```
~
:set fileformat=unix
```

mapper, reducer 실행권한 부여 / 샘플 파일 생성

```
[root@hadoop101 MapReduce]# chmod 755 word_mapper.py
[root@hadoop101 MapReduce]# chmod +x  word_reducer.py
```

mapper, reducer
실행권한 부여

```
[root@hadoop101 MapReduce]# vi sample.txt
      1 Deer Bear River
      2 Car Car River
      3 Deer Car Bear
```

샘플 파일 생성

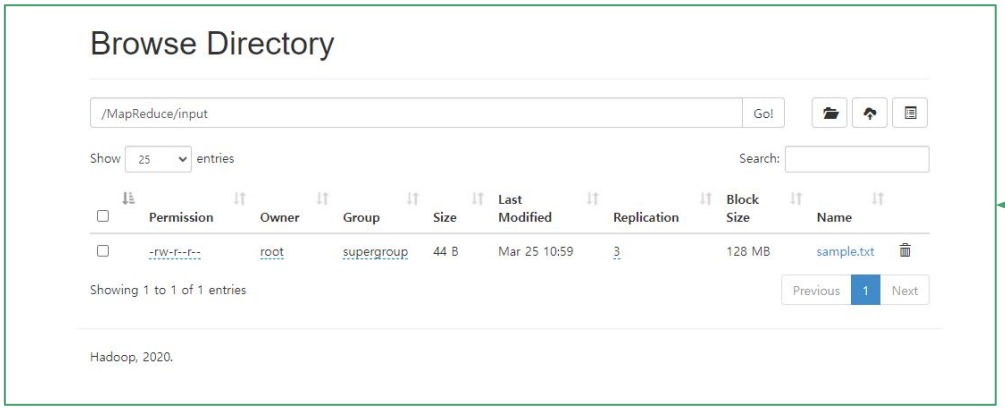
mapper, reducer 테스트 / HDFS MapReduce 디렉터리 생성

```
[root@hadoop101 MapReduce]# cat sample.txt | ./word_mapper.py | sort | ./word_reducer.py
Bear    2
Car     3
Deer    2
River   2
```

mapper, reducer 테스트

```
[root@hadoop101 MapReduce]# hdfs dfs -mkdir  /MapReduce
mkdir: `/MapReduce': File exists
[root@hadoop101 MapReduce]# hdfs dfs -mkdir  /MapReduce/input
[root@hadoop101 MapReduce]# hdfs dfs -put  sample.txt   /MapReduce/input
[root@hadoop101 MapReduce]#
```

HDFS MapReduce 디렉터리 생성
sample.txt put



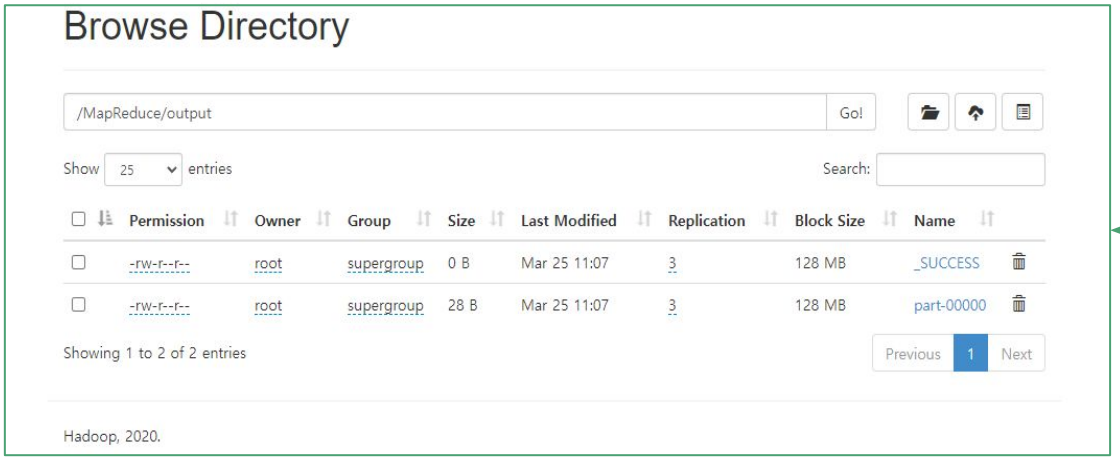
HDFS MapReduce 디렉터리 생성
sample.txt put 확인

Hadoop MapReduce 실행

```
[root@hadoop101 MapReduce]# hadoop  jar  hadoop-streaming-2.10.1.jar \
> -input  /MapReduce/input/sample.txt \
> -output  /MapReduce/output \
> -mapper word_mapper.py \
> -reducer word_reducer.py \
> -file /root/MapReduce/word_mapper.py \
> -file /root/MapReduce/word_reducer.py
```

```
22/03/25 11:07:04 INFO mapreduce.Job:  map 0% reduce 0%
22/03/25 11:07:13 INFO mapreduce.Job:  map 100% reduce 0%
22/03/25 11:07:18 INFO mapreduce.Job:  map 100% reduce 100%
22/03/25 11:07:18 INFO mapreduce.Job: Job job_1648173421667_0001 completed successfully
22/03/25 11:07:18 INFO mapreduce.Job: Counters: 49
```

Hadoop MapReduce 실행 확인



output 디렉터리확인

```
[root@hadoop101 MapReduce]# hdfs dfs -get /MapReduce/output/part-00000
[root@hadoop101 MapReduce]# ll
합 계  148
-rw-r--r-- 1 root root 135107  3월  25 10:39 hadoop-streaming-2.10.1.jar
-rw-r--r-- 1 root root     28  3월  25 11:09 part-00000
-rw-r--r-- 1 root root     44  3월  25 10:54 sample.txt
-rwxr-xr-x 1 root root    151  3월  25 10:45 word_mapper.py
-rwxr-xr-x 1 root root    276  3월  25 10:52 word_reducer.py
[root@hadoop101 MapReduce]# cat part-00000
Bear    2
Car     3
Deer    2
River   2
```

get해서 파일 확인