

Will Townsend

COSC 420

Homework 03

This assignment was pretty straight forward. It was interesting seeing the power of MapReduce. Below are the commands used with the terminal output. And attached are the WordCount.java, output, and the two test files used.

```
wtownsend2@hpc15-4:~/COS420/Homework/hw04$ hadoop com.sun.tools.javac.Main WordCount.java
```

```
wtownsend2@hpc15-4:~/COS420/Homework/hw04$ jar cf wc.jar WordCount*.class
```

```
wtownsend2@hpc15-4:~/COS420/Homework/hw04$ hdfs dfs -ls /user/wtownsend2/wordcount
```

Found 1 items

```
drwxr-xr-x  - wtownsend2 supergroup      0 2022-10-28 10:45 /user/wtownsend2/wordcount/input
```

```
wtownsend2@hpc15-4:~/COS420/Homework/hw04$ hadoop jar wc.jar WordCount
/user/wtownsend2/wordcount/input /user/wtownsend2/wordcount/output
```

```
2022-11-03 10:41:16,165 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at
hpc16-1.hslinux/10.249.43.32:8032
```

```
2022-11-03 10:41:16,503 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
```

```
2022-11-03 10:41:16,532 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path:
/user/wtownsend2/.staging/job_1667234420360_0004
```

```
2022-11-03 10:41:16,838 INFO input.FileInputFormat: Total input files to process : 4
```

```
2022-11-03 10:41:17,047 INFO mapreduce.JobSubmitter: number of splits:4
```

```
2022-11-03 10:41:17,205 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667234420360_0004
```

```
2022-11-03 10:41:17,205 INFO mapreduce.JobSubmitter: Executing with tokens: []
```

```
2022-11-03 10:41:17,389 INFO conf.Configuration: resource-types.xml not found
```

```
2022-11-03 10:41:17,389 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
```

```
2022-11-03 10:41:17,457 INFO impl.YarnClientImpl: Submitted application application_1667234420360_0004
```

```
2022-11-03 10:41:17,497 INFO mapreduce.Job: The url to track the job: http://hpc16-
1.hslinux:8088/proxy/application_1667234420360_0004/
```

```
2022-11-03 10:41:17,498 INFO mapreduce.Job: Running job: job_1667234420360_0004
```

```
2022-11-03 10:41:22,640 INFO mapreduce.Job: Job job_1667234420360_0004 running in uber mode : false
```

```
2022-11-03 10:41:22,642 INFO mapreduce.Job:  map 0% reduce 0%
```

```
2022-11-03 10:41:28,759 INFO mapreduce.Job:  map 100% reduce 0%
```

```
2022-11-03 10:41:33,799 INFO mapreduce.Job:  map 100% reduce 100%
```

```
2022-11-03 10:41:33,811 INFO mapreduce.Job: Job job_1667234420360_0004 completed successfully
```

```
2022-11-03 10:41:33,926 INFO mapreduce.Job: Counters: 56
```

File System Counters

FILE: Number of bytes read=345512

FILE: Number of bytes written=2069962

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=934748

HDFS: Number of bytes written=121681

HDFS: Number of read operations=17

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed map tasks=1

Launched map tasks=4

Launched reduce tasks=1

Data-local map tasks=1

Rack-local map tasks=3

Total time spent by all maps in occupied slots (ms)=13435

Total time spent by all reduces in occupied slots (ms)=2583

Total time spent by all map tasks (ms)=13435

Total time spent by all reduce tasks (ms)=2583

Total vcore-milliseconds taken by all map tasks=13435

Total vcore-milliseconds taken by all reduce tasks=2583

Total megabyte-milliseconds taken by all map tasks=13757440

Total megabyte-milliseconds taken by all reduce tasks=2644992

Map-Reduce Framework

Map input records=21980

Map output records=163424

Map output bytes=1560184

Map output materialized bytes=345530

Input split bytes=528

Combine input records=163424

Combine output records=25250

Reduce input groups=12254

Reduce shuffle bytes=345530

Reduce input records=25250

Reduce output records=12254

Spilled Records=50500

Shuffled Maps =4

Failed Shuffles=0

Merged Map outputs=4

GC time elapsed (ms)=511

CPU time spent (ms)=12220

Physical memory (bytes) snapshot=2265837568

Virtual memory (bytes) snapshot=13131825152

Total committed heap usage (bytes)=2805465088

Peak Map Physical memory (bytes)=621932544

Peak Map Virtual memory (bytes)=2628751360

Peak Reduce Physical memory (bytes)=291201024

Peak Reduce Virtual memory (bytes)=2635321344

Shuffle Errors

BAD_ID=0

CONNECTION=0

IO_ERROR=0

WRONG_LENGTH=0

WRONG_MAP=0

WRONG_REDUCE=0

File Input Format Counters

Bytes Read=934220

File Output Format Counters

Bytes Written=121681