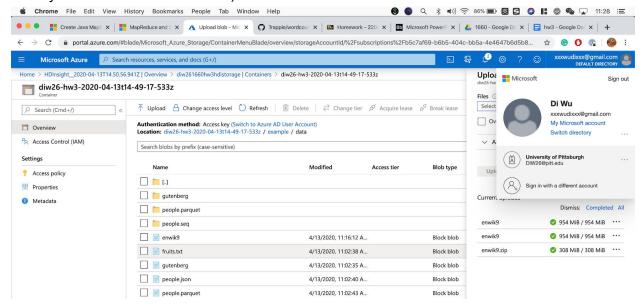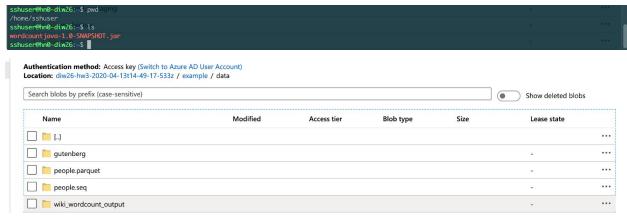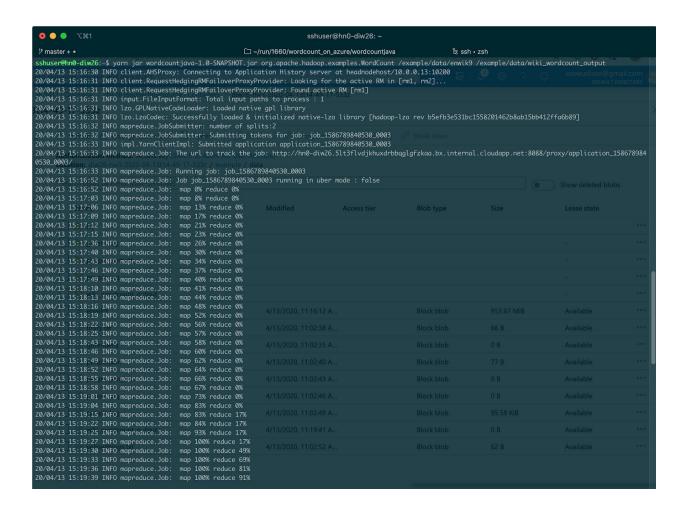Azure with username (I have used azure with my pitt email before, there is no free credits, so I use my own email address for this):
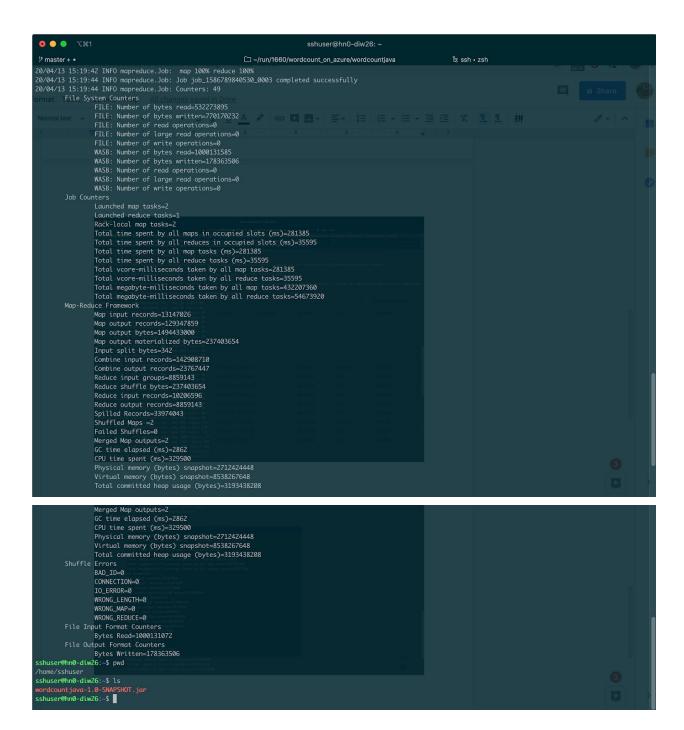

Word count implementation on azure (wiki_wordcount_output is the output folder)

Output:

sshuser@hn0-diw26:~$ yarn jar wordcountjava-1.0-SNAPSHOT.jar org.apache.hadoop.examples.WordCount /example/data/enwik9 /example/data/wiki_wordcount_output
20/04/13 15:16:30 INFO client.AHSProxy: Connecting to Application History server at headnodehost/10.0.0.13:10200
20/04/13 15:16:31 INFO client.RequestHedgingRMFailoverProxyProvider: Looking for the active RM in [rm1, rm2]...
20/04/13 15:16:31 INFO client.RequestHedgingRMFailoverProxyProvider: Found active RM [rm1]
20/04/13 15:16:31 INFO input.FileInputFormat: Total input paths to process : 1
20/04/13 15:16:31 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
20/04/13 15:16:31 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev b5efb3e531bc155820146Zb8ab15bb412ffa6b89]
20/04/13 15:16:32 INFO mapreduce.JobSubmitter: number of splits:2
20/04/13 15:16:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1586789840530_0003
20/04/13 15:16:33 INFO impl.YarnClientImpl: Submitted application application_1586789840530_0003
20/04/13 15:16:33 INFO mapreduce.Job: The url to track the job: http://hn0-diw26.5lt3flvdjkhuxdrbbqglgfzkaa.bx.internal.cloudapp.net:8088/proxy/application_158678984
0530_0003/
20/04/13 15:16:33 INFO mapreduce.Job: Running job: job_1586789840530_0003
20/04/13 15:16:52 INFO mapreduce.Job: Job job_1586789840530_0003 running in uber mode : false
20/04/13 15:16:52 INFO mapreduce.Job: map 0% reduce 0%
20/04/13 15:17:03 INFO mapreduce.Job: map 8% reduce 0%
20/04/13 15:17:06 INFO mapreduce.Job: map 13% reduce 0%
20/04/13 15:17:09 INFO mapreduce.Job: map 17% reduce 0%
20/04/13 15:17:12 INFO mapreduce.Job: map 21% reduce 0%
20/04/13 15:17:15 INFO mapreduce.Job: map 23% reduce 0%
20/04/13 15:17:36 INFO mapreduce.Job: map 26% reduce 0%
20/04/13 15:17:40 INFO mapreduce.Job: map 30% reduce 0%
20/04/13 15:17:43 INFO mapreduce.Job: map 34% reduce 0%
20/04/13 15:17:46 INFO mapreduce.Job: map 37% reduce 0%
20/04/13 15:17:49 INFO mapreduce.Job: map 40% reduce 0%
20/04/13 15:18:10 INFO mapreduce.Job: map 41% reduce 0%
20/04/13 15:18:13 INFO mapreduce.Job: map 44% reduce 0%
20/04/13 15:18:16 INFO mapreduce.Job: map 48% reduce 0%
20/04/13 15:18:19 INFO mapreduce.Job: map 52% reduce 0%
20/04/13 15:18:22 INFO mapreduce.Job: map 56% reduce 0%
20/04/13 15:18:25 INFO mapreduce.Job: map 57% reduce 0%
20/04/13 15:18:43 INFO mapreduce.Job: map 58% reduce 0%
20/04/13 15:18:46 INFO mapreduce.Job: map 60% reduce 0%
20/04/13 15:18:49 INFO mapreduce.Job: map 62% reduce 0%
20/04/13 15:18:52 INFO mapreduce.Job: map 64% reduce 0%
20/04/13 15:18:55 INFO mapreduce.Job: map 66% reduce 0%
20/04/13 15:18:58 INFO mapreduce.Job: map 67% reduce 0%
20/04/13 15:19:01 INFO mapreduce.Job: map 73% reduce 0%
20/04/13 15:19:04 INFO mapreduce.Job: map 83% reduce 0%
20/04/13 15:19:15 INFO mapreduce.Job: map 83% reduce 17%
20/04/13 15:19:22 INFO mapreduce.Job: map 84% reduce 17%
20/04/13 15:19:25 INFO mapreduce.Job: map 93% reduce 17%
20/04/13 15:19:27 INFO mapreduce.Job: map 100% reduce 17%
20/04/13 15:19:30 INFO mapreduce.Job: map 100% reduce 49%
20/04/13 15:19:33 INFO mapreduce.Job: map 100% reduce 69%
20/04/13 15:19:36 INFO mapreduce.Job: map 100% reduce 81%
20/04/13 15:19:39 INFO mapreduce.Job: map 100% reduce 91%

Modified	Access tier	Blob type	Size	Lease state	
					...
				-	...
				-	...
				-	...
				-	...
4/13/2020, 11:16:12 A...		Block blob	953.67 MiB	Available	...
4/13/2020, 11:02:38 A...		Block blob	66 B	Available	...
4/13/2020, 11:02:35 A...		Block blob	0 B	Available	...
4/13/2020, 11:02:40 A...		Block blob	77 B	Available	...
4/13/2020, 11:02:43 A...		Block blob	0 B	Available	...
4/13/2020, 11:02:46 A...		Block blob	0 B	Available	...
4/13/2020, 11:02:49 A...		Block blob	95.59 KiB	Available	...
4/13/2020, 11:19:41 A...		Block blob	0 B	Available	...
4/13/2020, 11:02:52 A...		Block blob	62 B	Available	...

```
20/04/13 15:19:42 INFO mapreduce.Job:  map 100% reduce 100%
20/04/13 15:19:44 INFO mapreduce.Job: Job job_1586789840530_0003 completed successfully
20/04/13 15:19:44 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=532273895
                FILE: Number of bytes written=770170232
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                WASB: Number of bytes read=1000131585
                WASB: Number of bytes written=178363506
                WASB: Number of read operations=0
                WASB: Number of large read operations=0
                WASB: Number of write operations=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Rack-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=281385
                Total time spent by all reduces in occupied slots (ms)=35595
                Total time spent by all map tasks (ms)=281385
                Total time spent by all reduce tasks (ms)=35595
                Total vcore-milliseconds taken by all map tasks=281385
                Total vcore-milliseconds taken by all reduce tasks=35595
                Total megabyte-milliseconds taken by all map tasks=432207360
                Total megabyte-milliseconds taken by all reduce tasks=54673920
        Map-Reduce Framework
                Map input records=13147026
                Map output records=129347859
                Map output bytes=1494433000
                Map output materialized bytes=237403654
                Input split bytes=342
                Combine input records=142908710
                Combine output records=23767447
                Reduce input groups=8859143
                Reduce shuffle bytes=237403654
                Reduce input records=10206596
                Reduce output records=8859143
                Spilled Records=33974043
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=2862
                CPU time spent (ms)=329500
                Physical memory (bytes) snapshot=2712424448
                Virtual memory (bytes) snapshot=8538267648
                Total committed heap usage (bytes)=3193438208
                Merged Map outputs=2
                GC time elapsed (ms)=2862
                CPU time spent (ms)=329500
                Physical memory (bytes) snapshot=2712424448
                Virtual memory (bytes) snapshot=8538267648
                Total committed heap usage (bytes)=3193438208
        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=1000131072
        File Output Format Counters
                Bytes Written=178363506
sshuser@hn0-diw26:~$ pwd
/home/sshuser
sshuser@hn0-diw26:~$ ls
wordcountjava-1.0-SNAPSHOT.jar
sshuser@hn0-diw26:~$
```

```
Luleå    15
Luleå''  1
Luleå''.        1
Luleå's  1
Luleå,   1
Luleå]   1
Luleå]]  2
Lulgiurai]]      1
Luli     7
Luli's   1
Luli]]   1
Luling   12
Luling's         1
Luling,  3
Luling.  2
Luling's         2
Lull     9
Lull&quot;,      1
Lull'd   1
Lull),   1
Lull,    1
Lull</title>     3
Lull]]   1
Lull]]&lt;small&gt;&lt;sup&gt;1&lt;/sup&gt;&lt;/small&gt;        1
Lull]]'s         1
Lull]],  2
Lulla    5
Lullabelle       1
Lullabies        4
Lullabies''      1
```

The link of input and output file (input is 1G, output is 178M)

Input:

https://pitt-my.sharepoint.com/:u:/g/personal/diw26_pitt_edu/EU5-Gp_gDyxFgS4mEqcb1NcBDsyzMM_RqaAho_PQc5fmsA?e=Y8vNqz

Output:

https://pitt-my.sharepoint.com/:u:/g/personal/diw26_pitt_edu/EScmtNz4ydpKgzClnWFw5REB3iNTti-udSesctcJRx1I_g?e=P8wXLl