```
Last login: Sat Sep 29 23:47:07 on ttys001
Noopurs-MacBook-Pro:Desktop noopurrkalawatia$ ssh -i hadoop-word-count.pem hadoop@ec2-18-188-97-172.us-east-2.compute.amazonaws.co
m
The authenticity of host 'ec2-18-188-97-172.us-east-2.compute.amazonaws.com (18.188.97.172)' can't be established.
ECDSA key fingerprint is SHA256:qWaE0Jh4l0G2IuQV3ZJEDpddXBwNonYuDIu7A4UF95g.
Are you sure you want to continue connecting (yes/no)? yes
Failed to add the host to the list of known hosts (/Users/noopurrkalawatia/.ssh/known_hosts).
Last login: Sun Sep 30 05:32:22 2018

       __|  __|_  )
       _|  (     /   Amazon Linux AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
14 package(s) needed for security, out of 30 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE::::EEEEEEEEE::::E M::::::::M         M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M       M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M     M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE    M:::::M M:::M M:::M M:::::M    R:::RRRRRR:::::R
  E::::::::::::::E    M:::::M  M:::M:::M  M:::::M    R:::::::::::RR
  E:::::EEEEEEEEEE    M:::::M   M:::::M   M:::::M    R:::RRRRRR:::::R
  E::::E             M:::::M    M:::M    M:::::M    R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M    R:::R      R::::R
EE::::EEEEEEEE::::E M:::::M             M:::::M    R:::R      R::::R
E::::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR


[hadoop@ip-172-31-5-137 ~]$ ls
wordcount
[hadoop@ip-172-31-5-137 ~]$ cd wordcount
[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount3.jar com.wordcount3.WordCount s3://biblewordcount/ s3://biblewordcount/ou
tput2/
18/09/30 05:41:48 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:80
32
18/09/30 05:41:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interfa
ce and execute your application with ToolRunner to remedy this.
18/09/30 05:41:50 INFO input.FileInputFormat: Total input files to process : 10
18/09/30 05:41:50 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 05:41:50 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a3
2e5e2d9e5dfd7724]
18/09/30 05:41:50 INFO mapreduce.JobSubmitter: number of splits:10
18/09/30 05:41:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0004
18/09/30 05:41:50 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0004
18/09/30 05:41:51 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/appl
ication_1538278733852_0004/
18/09/30 05:41:51 INFO mapreduce.Job: Running job: job_1538278733852_0004
18/09/30 05:42:01 INFO mapreduce.Job: Job job_1538278733852_0004 running in uber mode : false
18/09/30 05:42:01 INFO mapreduce.Job:  map 0% reduce 0%
18/09/30 05:42:17 INFO mapreduce.Job:  map 10% reduce 0%
18/09/30 05:42:18 INFO mapreduce.Job:  map 20% reduce 0%
18/09/30 05:42:19 INFO mapreduce.Job:  map 40% reduce 0%
18/09/30 05:42:23 INFO mapreduce.Job:  map 60% reduce 0%
18/09/30 05:42:25 INFO mapreduce.Job:  map 80% reduce 0%
18/09/30 05:42:26 INFO mapreduce.Job:  map 100% reduce 100%
18/09/30 05:42:27 INFO mapreduce.Job: Job job_1538278733852_0004 completed successfully
18/09/30 05:42:27 INFO mapreduce.Job: Counters: 55
        File System Counters
                FILE: Number of bytes read=1745
                FILE: Number of bytes written=1884895
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=985
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
                S3: Number of bytes read=81613300
                S3: Number of bytes written=891
                S3: Number of read operations=0
                S3: Number of large read operations=0
                S3: Number of write operations=0
        Job Counters
                Killed map tasks=1
                Launched map tasks=10
                Launched reduce tasks=1
                Data-local map tasks=10
                Total time spent by all maps in occupied slots (ms)=9023808
                Total time spent by all reduces in occupied slots (ms)=675456
                Total time spent by all map tasks (ms)=187996
                Total time spent by all reduce tasks (ms)=7036
                Total vcore-milliseconds taken by all map tasks=187996
```

```
                Total vcore-milliseconds taken by all reduce tasks=7036
                Total megabyte-milliseconds taken by all map tasks=288761856
                Total megabyte-milliseconds taken by all reduce tasks=21614592
        Map-Reduce Framework
                Map input records=1405946
                Map output records=6067351
                Map output bytes=47130252
                Map output materialized bytes=8332
                Input split bytes=985
                Combine input records=6067351
                Combine output records=820
                Reduce input groups=82
                Reduce shuffle bytes=8332
                Reduce input records=820
                Reduce output records=82
                Spilled Records=1640
                Shuffled Maps =10
                Failed Shuffles=0
                Merged Map outputs=10
                GC time elapsed (ms)=7774
                CPU time spent (ms)=45100
                Physical memory (bytes) snapshot=7122997248
                Virtual memory (bytes) snapshot=37727354880
                Total committed heap usage (bytes)=6552027136
        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=81613300
        File Output Format Counters
                Bytes Written=891
[hadoop@ip-172-31-5-137 wordcount]$ rm WordCount3.jar
[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount3.jar com.wordcount3.WordCount s3://biblewordcount/ s3://biblewordcount/ou
tput2/
18/09/30 06:05:53 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:80
32
18/09/30 06:05:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interfa
ce and execute your application with ToolRunner to remedy this.
18/09/30 06:05:55 INFO input.FileInputFormat: Total input files to process : 10
18/09/30 06:05:56 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 06:05:56 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a3
2e5e2d9e5dfd7724]
18/09/30 06:05:56 INFO mapreduce.JobSubmitter: number of splits:10
18/09/30 06:05:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0005
18/09/30 06:05:56 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0005
18/09/30 06:05:56 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/appl
ication_1538278733852_0005/
18/09/30 06:05:56 INFO mapreduce.Job: Running job: job_1538278733852_0005
18/09/30 06:06:07 INFO mapreduce.Job: Job job_1538278733852_0005 running in uber mode : false
18/09/30 06:06:07 INFO mapreduce.Job:  map 0% reduce 0%
18/09/30 06:06:24 INFO mapreduce.Job:  map 30% reduce 0%
18/09/30 06:06:25 INFO mapreduce.Job:  map 40% reduce 0%
18/09/30 06:06:29 INFO mapreduce.Job:  map 50% reduce 0%
18/09/30 06:06:31 INFO mapreduce.Job:  map 70% reduce 0%
18/09/30 06:06:32 INFO mapreduce.Job:  map 80% reduce 0%
18/09/30 06:06:33 INFO mapreduce.Job:  map 100% reduce 0%
18/09/30 06:06:34 INFO mapreduce.Job:  map 100% reduce 25%
18/09/30 06:06:37 INFO mapreduce.Job:  map 100% reduce 50%
18/09/30 06:06:38 INFO mapreduce.Job:  map 100% reduce 75%
18/09/30 06:06:39 INFO mapreduce.Job:  map 100% reduce 100%
18/09/30 06:06:39 INFO mapreduce.Job: Job job_1538278733852_0005 completed successfully
18/09/30 06:06:39 INFO mapreduce.Job: Counters: 55
        File System Counters
                FILE: Number of bytes read=2023
                FILE: Number of bytes written=2398437
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=985
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=10
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
                S3: Number of bytes read=81613300
                S3: Number of bytes written=891
                S3: Number of read operations=0
                S3: Number of large read operations=0
                S3: Number of write operations=0
        Job Counters
                Killed map tasks=1
                Launched map tasks=10
                Launched reduce tasks=4
```

```
                Data-local map tasks=10
                Total time spent by all maps in occupied slots (ms)=9160848
                Total time spent by all reduces in occupied slots (ms)=3446592
                Total time spent by all map tasks (ms)=190851
                Total time spent by all reduce tasks (ms)=35902
                Total vcore-milliseconds taken by all map tasks=190851
                Total vcore-milliseconds taken by all reduce tasks=35902
                Total megabyte-milliseconds taken by all map tasks=293147136
                Total megabyte-milliseconds taken by all reduce tasks=110290944
        Map-Reduce Framework
                Map input records=1405946
                Map output records=6067351
                Map output bytes=47130252
                Map output materialized bytes=9684
                Input split bytes=985
                Combine input records=6067351
                Combine output records=820
                Reduce input groups=82
                Reduce shuffle bytes=9684
                Reduce input records=820
                Reduce output records=82
                Spilled Records=1640
                Shuffled Maps =40
                Failed Shuffles=0
                Merged Map outputs=40
                GC time elapsed (ms)=7781
                CPU time spent (ms)=47820
                Physical memory (bytes) snapshot=8210546688
                Virtual memory (bytes) snapshot=51722678272
                Total committed heap usage (bytes)=7659323392
        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=81613300
        File Output Format Counters
                Bytes Written=891
[hadoop@ip-172-31-5-137 wordcount]$
```