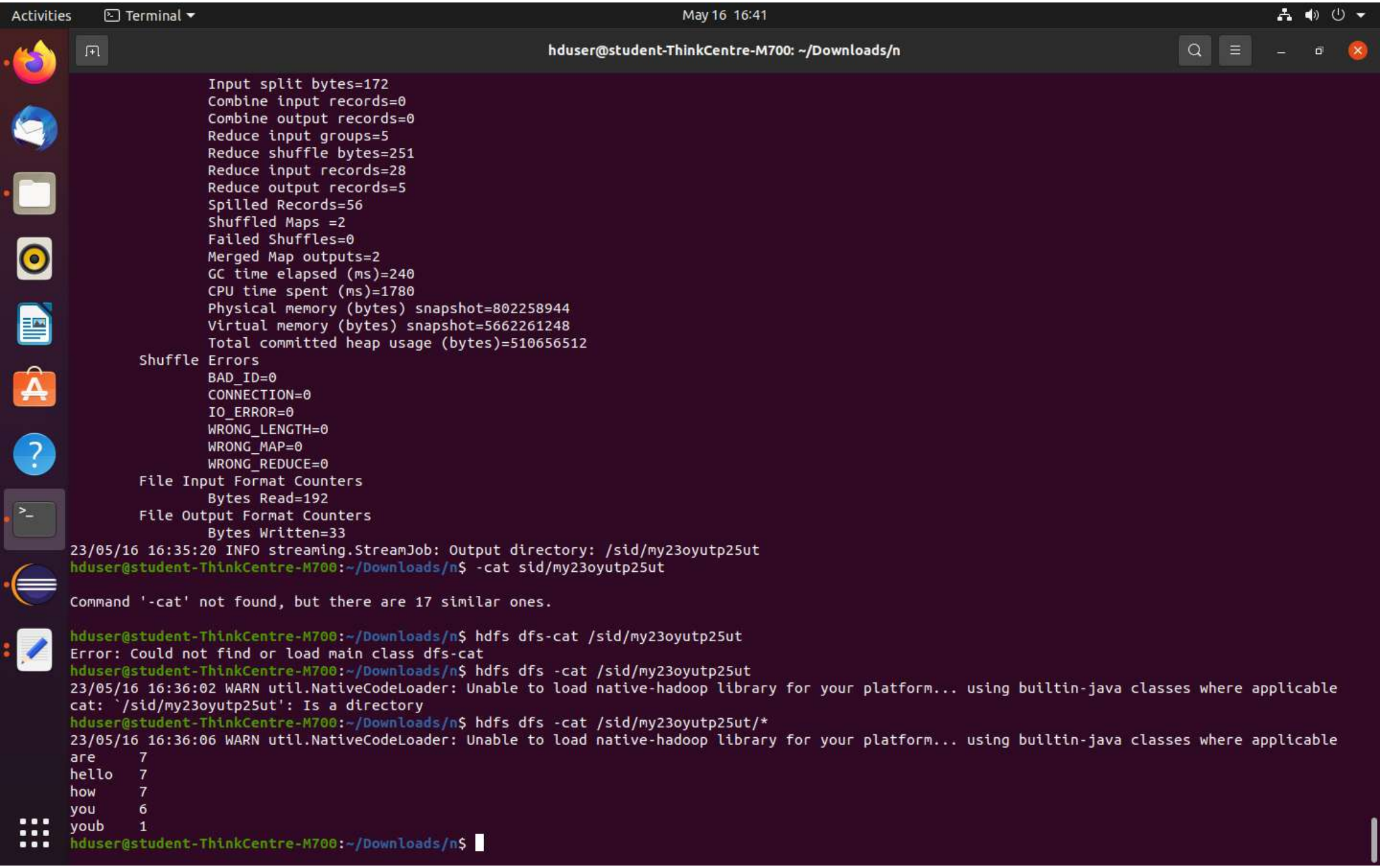
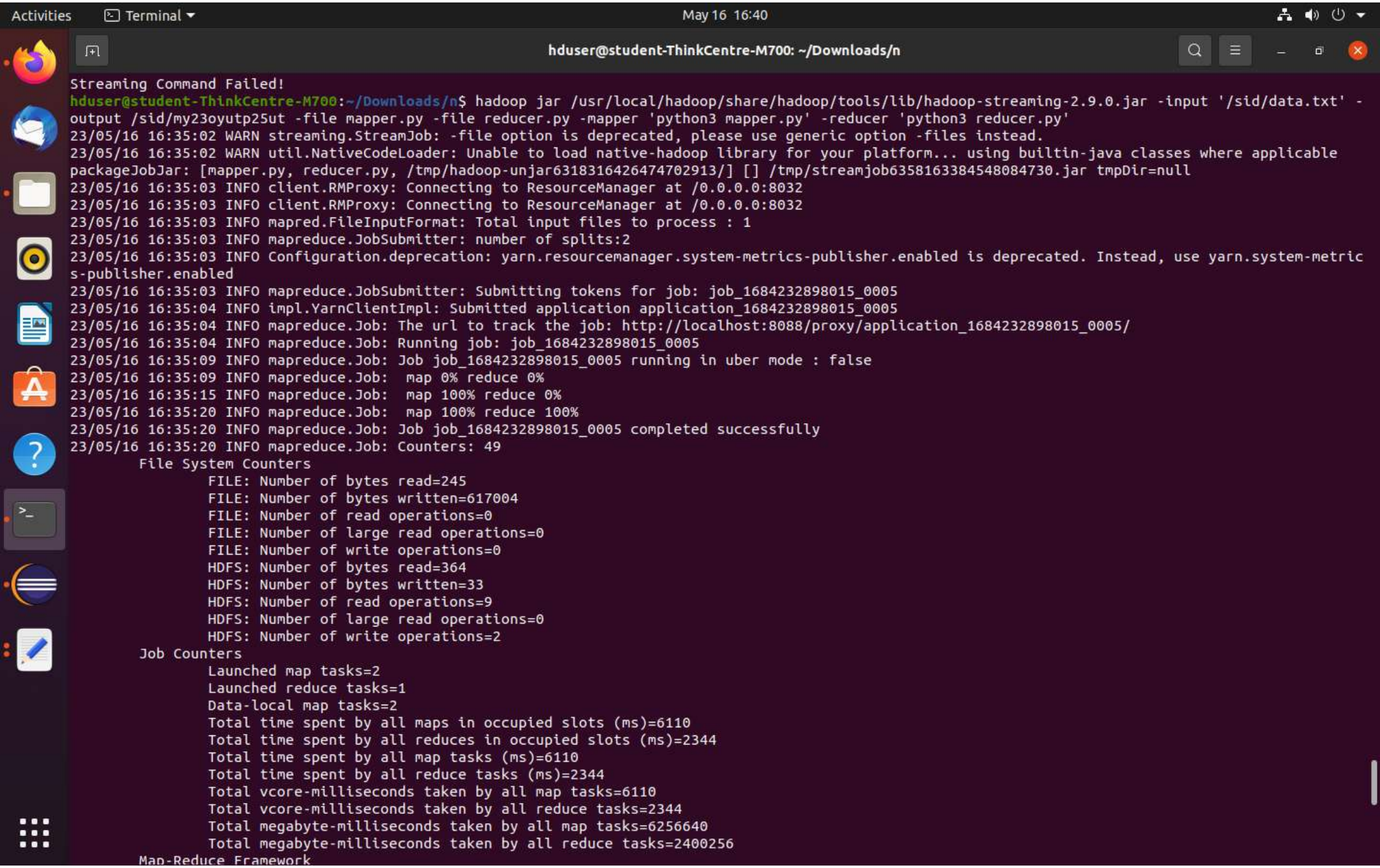
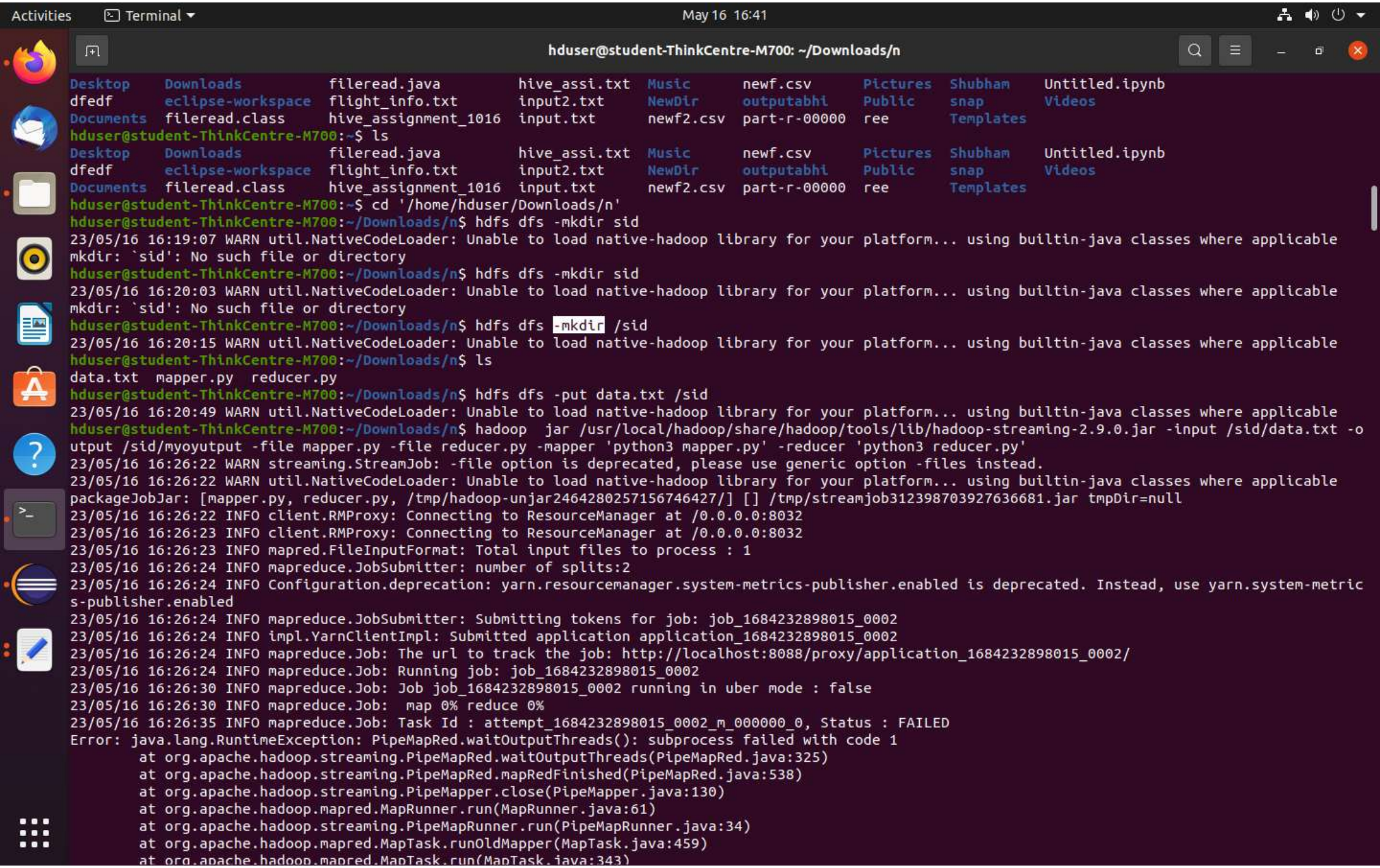
```
## Mapper

import sys
import io

input_stream = io.TextIOWrapper(sys.stdin.buffer)


for line in input_stream:
    line = line.lower()
    words = line.split()
    for word in words:
       print("%s\t%s" %(word,1))




## Reducer

import sys
import io

current_word = None
current_count = 0
word = None

for line in sys.stdin:
      word, count = line.split('\t',1)
      count = int(count)
      if current_word == word:
            current_count += count
      else:
            if current_word:
                  print("%s\t%s" %(current_word,current_count))
            current_count = count
            current_word = word

if current_word == word:
      print("%s\t%s" %(current_word,current_count))
```


Activities
Terminal
May 16 16:41
hduser@student-ThinkCentre-M700: ~/Downloads/n
Input split bytes=172
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=251
Reduce input records=28
Reduce output records=5
Spilled Records=56
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=240
CPU time spent (ms)=1780
Physical memory (bytes) snapshot=802258944
Virtual memory (bytes) snapshot=5662261248
Total committed heap usage (bytes)=510656512
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=192
File Output Format Counters
Bytes Written=33
23/05/16 16:35:20 INFO streaming.StreamJob: Output directory: /sid/my23oyutp25ut
hduser@student-ThinkCentre-M700:~/Downloads/n$ -cat sid/my23oyutp25ut
Command '-cat' not found, but there are 17 similar ones.
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs-cat /sid/my23oyutp25ut
Error: Could not find or load main class dfs-cat
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -cat /sid/my23oyutp25ut
23/05/16 16:36:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: `/sid/my23oyutp25ut': Is a directory
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -cat /sid/my23oyutp25ut/*
23/05/16 16:36:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
are 7
hello 7
how 7
you 6
youb 1
hduser@student-ThinkCentre-M700:~/Downloads/n$


Activities
Terminal
May 16 16:40
hduser@student-ThinkCentre-M700: ~/Downloads/n
Streaming Command Failed!
hduser@student-ThinkCentre-M700:~/Downloads/n$ hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.9.0.jar -input '/sid/data.txt' -output /sid/my23oyutp25ut -file mapper.py -file reducer.py -mapper 'python3 mapper.py' -reducer 'python3 reducer.py'
23/05/16 16:35:02 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
23/05/16 16:35:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
packageJobJar: [mapper.py, reducer.py, /tmp/hadoop-unjar6318316426474702913/] [] /tmp/streamjob6358163384548084730.jar tmpDir=null
23/05/16 16:35:03 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/05/16 16:35:03 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/05/16 16:35:03 INFO mapred.FileInputFormat: Total input files to process : 1
23/05/16 16:35:03 INFO mapreduce.JobSubmitter: number of splits:2
23/05/16 16:35:03 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/05/16 16:35:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684232898015_0005
23/05/16 16:35:04 INFO impl.YarnClientImpl: Submitted application application_1684232898015_0005
23/05/16 16:35:04 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1684232898015_0005/
23/05/16 16:35:04 INFO mapreduce.Job: Running job: job_1684232898015_0005
23/05/16 16:35:09 INFO mapreduce.Job: Job job_1684232898015_0005 running in uber mode : false
23/05/16 16:35:09 INFO mapreduce.Job: map 0% reduce 0%
23/05/16 16:35:15 INFO mapreduce.Job: map 100% reduce 0%
23/05/16 16:35:20 INFO mapreduce.Job: map 100% reduce 100%
23/05/16 16:35:20 INFO mapreduce.Job: Job job_1684232898015_0005 completed successfully
23/05/16 16:35:20 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=245
FILE: Number of bytes written=617004
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=364
HDFS: Number of bytes written=33
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=6110
Total time spent by all reduces in occupied slots (ms)=2344
Total time spent by all map tasks (ms)=6110
Total time spent by all reduce tasks (ms)=2344
Total vcore-milliseconds taken by all map tasks=6110
Total vcore-milliseconds taken by all reduce tasks=2344
Total megabyte-milliseconds taken by all map tasks=6256640
Total megabyte-milliseconds taken by all reduce tasks=2400256
Map-Reduce Framework


Activities
Terminal
May 16 16:41
hduser@student-ThinkCentre-M700: ~/Downloads/n
Desktop Downloads fileread.java hive_assi.txt Music newf.csv Pictures Shubham Untitled.ipynb
dfedf eclipse-workspace flight_info.txt input2.txt NewDir outputabhi Public snap Videos
Documents fileread.class hive_assignment_1016 input.txt newf2.csv part-r-00000 ree Templates
hduser@student-ThinkCentre-M700:~$ ls
Desktop Downloads fileread.java hive_assi.txt Music newf.csv Pictures Shubham Untitled.ipynb
dfedf eclipse-workspace flight_info.txt input2.txt NewDir outputabhi Public snap Videos
Documents fileread.class hive_assignment_1016 input.txt newf2.csv part-r-00000 ree Templates
hduser@student-ThinkCentre-M700:~$ cd '/home/hduser/Downloads/n'
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -mkdir sid
23/05/16 16:19:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: `sid': No such file or directory
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -mkdir sid
23/05/16 16:20:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: `sid': No such file or directory
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -mkdir /sid
23/05/16 16:20:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@student-ThinkCentre-M700:~/Downloads/n$ ls
data.txt mapper.py reducer.py
hduser@student-ThinkCentre-M700:~/Downloads/n$ hdfs dfs -put data.txt /sid
23/05/16 16:20:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@student-ThinkCentre-M700:~/Downloads/n$ hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.9.0.jar -input /sid/data.txt -output /sid/myoyutput -file mapper.py -file reducer.py -mapper 'python3 mapper.py' -reducer 'python3 reducer.py'
23/05/16 16:26:22 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
23/05/16 16:26:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
packageJobJar: [mapper.py, reducer.py, /tmp/hadoop-unjar2464280257156746427/] [] /tmp/streamjob312398703927636681.jar tmpDir=null
23/05/16 16:26:22 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/05/16 16:26:23 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/05/16 16:26:23 INFO mapred.FileInputFormat: Total input files to process : 1
23/05/16 16:26:24 INFO mapreduce.JobSubmitter: number of splits:2
23/05/16 16:26:24 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
23/05/16 16:26:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684232898015_0002
23/05/16 16:26:24 INFO impl.YarnClientImpl: Submitted application application_1684232898015_0002
23/05/16 16:26:24 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1684232898015_0002/
23/05/16 16:26:24 INFO mapreduce.Job: Running job: job_1684232898015_0002
23/05/16 16:26:30 INFO mapreduce.Job: Job job_1684232898015_0002 running in uber mode : false
23/05/16 16:26:30 INFO mapreduce.Job: map 0% reduce 0%
23/05/16 16:26:35 INFO mapreduce.Job: Task Id : attempt_1684232898015_0002_m_000000_0, Status : FAILED
Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 1
at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:325)
at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:538)
at org.apache.hadoop.streaming.PipeMapper.close(PipeMapper.java:130)
at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:61)
at org.apache.hadoop.streaming.PipeMapRunner.run(PipeMapRunner.java:34)
at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:459)
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)