reducer.py

```python
import sys

# Initialize variables
temperature_sum = 0
temperature_count = 0
dewpoint_sum = 0
dewpoint_count = 0
windspeed_sum = 0
windspeed_count = 0

# Read input from standard input
for line in sys.stdin:
    # Remove leading and trailing whitespace
    line = line.strip()

    # Split the line into key and value
    key, value = line.split('\t', 1)

    # Update sums and counts based on key
    if key == 'temperature':
        temperature_sum += float(value)
        temperature_count += 1
    elif key == 'dewpoint':
        dewpoint_sum += float(value)
        dewpoint_count += 1
    elif key == 'windspeed':
        windspeed_sum += float(value)
        windspeed_count += 1

# Calculate averages
average_temperature = temperature_sum / temperature_count if temperature_count > 0 else 0
average_dewpoint = dewpoint_sum / dewpoint_count if dewpoint_count > 0 else 0
average_windspeed = windspeed_sum / windspeed_count if windspeed_count > 0 else 0

# Output the averages
print('Average Temperature:', average_temperature)
print('Average Dew Point:', average_dewpoint)
print('Average Wind Speed:', average_windspeed)
```

mapper.py

```python
import sys

# Initialize variables
temperature_sum = 0
temperature_count = 0
```

```python
dewpoint_sum = 0
dewpoint_count = 0
windspeed_sum = 0
windspeed_count = 0

# Read input from standard input
for line in sys.stdin:
    # Remove leading and trailing whitespace
    line = line.strip()

    # Skip empty lines
    if not line:
        continue

    # Split the line into temperature, dew point, and wind speed
    values = line.split(',')

    # Check if all values are present
    if len(values) != 3:
        continue

    try:
        temperature, dewpoint, windspeed = map(float, values)

        # Output key-value pairs
        print('temperature\t{}'.format(temperature))
        print('dewpoint\t{}'.format(dewpoint))
        print('windspeed\t{}'.format(windspeed))
    except ValueError:
        # Handle the case where conversion to float fails
        continue
```

Terminal:

aiml@aiml-System-Product-Name:~$ cd Desktop/reduce2

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Traceback (most recent call last):

  File "/home/aiml/Desktop/reduce2/mapper.py", line 18, in <module>

    count = int(values_count[1])

IndexError: list index out of range

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ ^C

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Average Temperature: 0

Average Dew Point: 0

Average Wind Speed: 0

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Traceback (most recent call last):

  File "/home/aiml/Desktop/reduce2/mapper.py", line 17, in <module>

    temperature, dewpoint, windspeed = map(float, line.split(','))

ValueError: could not convert string to float: ''

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

temperature_sum: 129.1

temperature_count: 6

dewpoint_sum: 81.39999999999999

dewpoint_count: 6

windspeed_sum: 35.7

windspeed_count: 6

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Traceback (most recent call last):

  File "/home/aiml/Desktop/reduce2/mapper.py", line 17, in <module>

    temperature, dewpoint, windspeed = map(float, line.split(','))

ValueError: could not convert string to float: ''

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Traceback (most recent call last):

  File "/home/aiml/Desktop/reduce2/mapper.py", line 17, in <module>

    temperature, dewpoint, windspeed = map(float, line.split(','))

ValueError: could not convert string to float: ''

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

Average Temperature: 21.516666666666666

Average Dew Point: 13.566666666666665

Average Wind Speed: 5.95

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat word_count_data.txt | python3 mapper.py | sort -k1,1 | python3 reducer.py

cat: word_count_data.txt: No such file or directory

Traceback (most recent call last):

  File "/home/aiml/Desktop/reduce2/reducer.py", line 16, in <module>

    key, value = line.split('\t', 1)

ValueError: not enough values to unpack (expected 2, got 1)

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py

temperature   23.4

dewpoint   15.2

windspeed   5.6

temperature   20.1

dewpoint   12.5

windspeed   6.3

temperature   18.5

dewpoint   10.8

windspeed   7.2

temperature   25.0

dewpoint   16.9

windspeed   4.8

temperature   22.3

dewpoint   14.7

windspeed   5.1

temperature   19.8

dewpoint   11.3

windspeed   6.7

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat word_count_data.txt | python3 mapper.py | sort -k1,1 | python3 reducer.py

cat: word_count_data.txt: No such file or directory

Average Temperature: 0

Average Dew Point: 0

Average Wind Speed: 0

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cat sample_weather.txt | python3 mapper.py | sort -k1,1 | python3 reducer.py

Average Temperature: 21.516666666666666

Average Dew Point: 13.566666666666668

Average Wind Speed: 5.949999999999999

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cd

aiml@aiml-System-Product-Name:~$ start-dfs.sh

Starting namenodes on [localhost]

localhost: namenode is running as process 28316. Stop it first and ensure /tmp/hadoop-aiml-namenode.pid file is empty before retry.

Starting datanodes

localhost: datanode is running as process 28494. Stop it first and ensure /tmp/hadoop-aiml-datanode.pid file is empty before retry.

Starting secondary namenodes [aiml-System-Product-Name]

aiml-System-Product-Name: secondarynamenode is running as process 28720. Stop it first and ensure /tmp/hadoop-aiml-secondarynamenode.pid file is empty before retry.

2024-04-04 11:47:12,356 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

aiml@aiml-System-Product-Name:~$ start-yarn.sh

Starting resourcemanager

resourcemanager is running as process 28950. Stop it first and ensure /tmp/hadoop-aiml-resourcemanager.pid file is empty before retry.

Starting nodemanagers

localhost: nodemanager is running as process 29291. Stop it first and ensure /tmp/hadoop-aiml-nodemanager.pid file is empty before retry.

aiml@aiml-System-Product-Name:~$ hdfs dfs -mkdir /avg_temp_in_python

2024-04-04 11:47:53,637 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

aiml@aiml-System-Product-Name:~$ hdfs dfs -mkdir /avg_temp_in_python

2024-04-04 11:47:58,538 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

mkdir: `/avg_temp_in_python': File exists

aiml@aiml-System-Product-Name:~$ hdfs dfs -copyFromLocal /home/aiml/Desktop/reduce2/sample_weather.txt /avg_temp_in_python

2024-04-04 11:48:45,619 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

aiml@aiml-System-Product-Name:~$ hdfs dfs -ls /

2024-04-04 11:49:00,561 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 10 items

```
drwxr-xr-x   - aiml supergroup          0 2024-04-04 11:48 /avg_temp_in_python
-rw-r--r--   1 aiml supergroup        399 2024-03-15 10:57 /data
-rw-r--r--   1 aiml supergroup        399 2024-03-15 10:57 /data.txt
drwxr-xr-x   - aiml supergroup          0 2024-03-15 10:58 /tmp
drwxr-xr-x   - aiml supergroup          0 2024-03-15 10:58 /user
drwxr-xr-x   - aiml supergroup          0 2024-04-04 10:06 /weather_forcast_in_python
drwxr-xr-x   - aiml supergroup          0 2024-04-03 15:45 /word_count_in_python
drwxr-xr-x   - aiml supergroup          0 2024-04-04 10:16 /word_count_in_python2
drwxr-xr-x   - aiml supergroup          0 2024-04-03 16:32 /word_count_in_python3
drwxr-xr-x   - aiml supergroup          0 2024-04-04 11:25 /word_count_in_python8
```

aiml@aiml-System-Product-Name:~$ hdfs dfs -ls /avg_temp_in_python

2024-04-04 11:49:34,286 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 1 items

```
-rw-r--r--   1 aiml supergroup         85 2024-04-04 11:48 /avg_temp_in_python/sample_weather.txt
```

aiml@aiml-System-Product-Name:~$ cd/Desktop/reduce2

bash: cd/Desktop/reduce2: No such file or directory

aiml@aiml-System-Product-Name:~$ cd/Desktop

bash: cd/Desktop: No such file or directory

aiml@aiml-System-Product-Name:~$ cd Desktop/reduce2

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ chmod 777 mapper.py reducer.py

aiml@aiml-System-Product-Name:~/Desktop/reduce2$ cd

aiml@aiml-System-Product-Name:~$ hadoop jar /home/aiml/Desktop/reduce2/hadoop-streaming-3.3.6.jar \

> -input /avg_temp_in_python/sample_weather.txt \

> -output /avg_temp_in_python/output \

> -mapper "python3 /home/aiml/Desktop/reduce2/mapper.py" \

> -reducer "python3 /home/aiml/Desktop/reduce2/reducer.py"

2024-04-04 11:53:18,540 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

packageJobJar: [/tmp/hadoop-unjar1566709327473520848/] []
/tmp/streamjob1191938198619220873.jar tmpDir=null

2024-04-04 11:53:18,872 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032

2024-04-04 11:53:18,948 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032

2024-04-04 11:53:19,058 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/aiml/.staging/job_1712204091628_0006

2024-04-04 11:53:19,189 INFO mapred.FileInputFormat: Total input files to process : 1

2024-04-04 11:53:19,211 INFO mapreduce.JobSubmitter: number of splits:2

2024-04-04 11:53:19,268 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1712204091628_0006

2024-04-04 11:53:19,268 INFO mapreduce.JobSubmitter: Executing with tokens: []

2024-04-04 11:53:19,364 INFO conf.Configuration: resource-types.xml not found

2024-04-04 11:53:19,364 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2024-04-04 11:53:19,401 INFO impl.YarnClientImpl: Submitted application application_1712204091628_0006

2024-04-04 11:53:19,422 INFO mapreduce.Job: The url to track the job: http://aiml-System-Product-Name:8088/proxy/application_1712204091628_0006/

2024-04-04 11:53:19,423 INFO mapreduce.Job: Running job: job_1712204091628_0006

2024-04-04 11:53:23,468 INFO mapreduce.Job: Job job_1712204091628_0006 running in uber mode : false

2024-04-04 11:53:23,470 INFO mapreduce.Job: map 0% reduce 0%

2024-04-04 11:53:26,520 INFO mapreduce.Job: map 100% reduce 0%

2024-04-04 11:53:30,554 INFO mapreduce.Job: map 100% reduce 100%

2024-04-04 11:53:30,568 INFO mapreduce.Job: Job job_1712204091628_0006 completed successfully

2024-04-04 11:53:30,621 INFO mapreduce.Job: Counters: 54

       File System Counters

              FILE: Number of bytes read=312

              FILE: Number of bytes written=834992

              FILE: Number of read operations=0

              FILE: Number of large read operations=0

              FILE: Number of write operations=0

HDFS: Number of bytes read=350

HDFS: Number of bytes written=106

HDFS: Number of read operations=11

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

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)=2710

Total time spent by all reduces in occupied slots (ms)=1156

Total time spent by all map tasks (ms)=2710

Total time spent by all reduce tasks (ms)=1156

Total vcore-milliseconds taken by all map tasks=2710

Total vcore-milliseconds taken by all reduce tasks=1156

Total megabyte-milliseconds taken by all map tasks=2775040

Total megabyte-milliseconds taken by all reduce tasks=1183744

Map-Reduce Framework

Map input records=7

Map output records=18

Map output bytes=270

Map output materialized bytes=318

Input split bytes=222

Combine input records=0

Combine output records=0

Reduce input groups=3

Reduce shuffle bytes=318

Reduce input records=18

Reduce output records=3

Spilled Records=36

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=89

CPU time spent (ms)=1010

Physical memory (bytes) snapshot=870002688

Virtual memory (bytes) snapshot=7719759872

Total committed heap usage (bytes)=931135488

Peak Map Physical memory (bytes)=329068544

Peak Map Virtual memory (bytes)=2571878400

Peak Reduce Physical memory (bytes)=216350720

Peak Reduce Virtual memory (bytes)=2577326080

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=128

File Output Format Counters

Bytes Written=106

2024-04-04 11:53:30,621 INFO streaming.StreamJob: Output directory: /avg_temp_in_python/output

aiml@aiml-System-Product-Name:~$ hdfs dfs -cat /avg_temp_in_python/output/part-00000

2024-04-04 11:54:05,549 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Average Temperature: 21.516666666666666

Average Dew Point: 13.566666666666668

Average Wind Speed: 5.95

aiml@aiml-System-Product-Name:~$

sample_weather.txt:

23.4,15.2,5.6

20.1,12.5,6.3

18.5,10.8,7.2

25.0,16.9,4.8

22.3,14.7,5.1

19.8,11.3,6.7