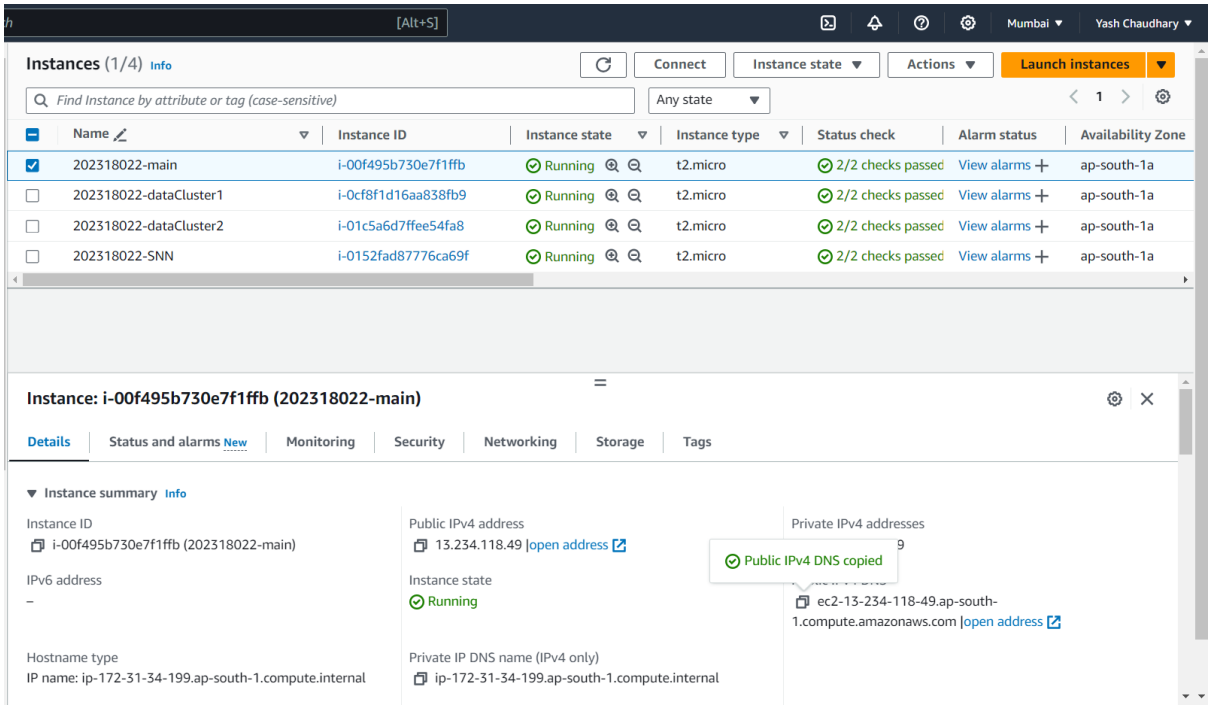


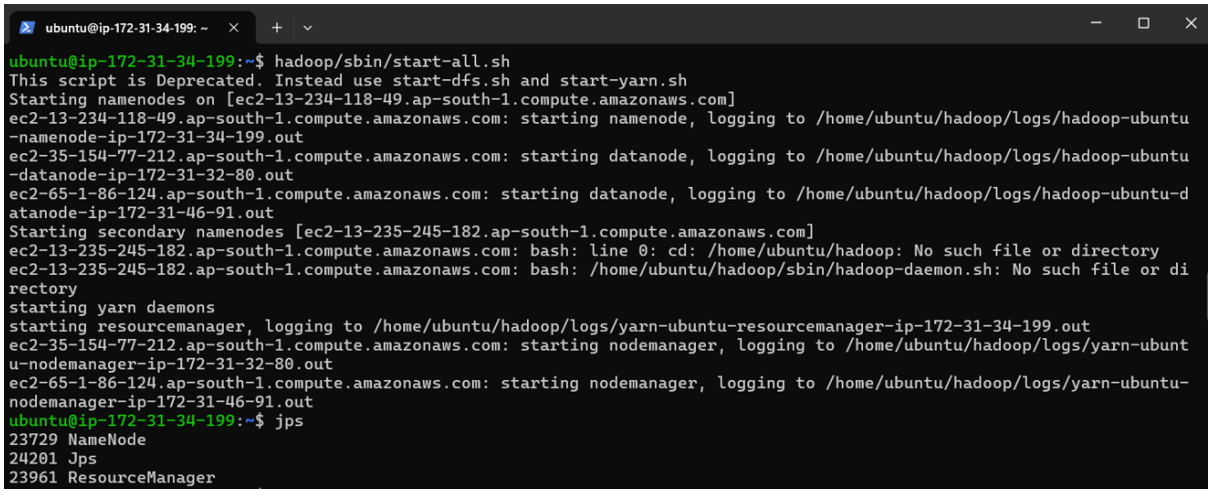
Mapper:

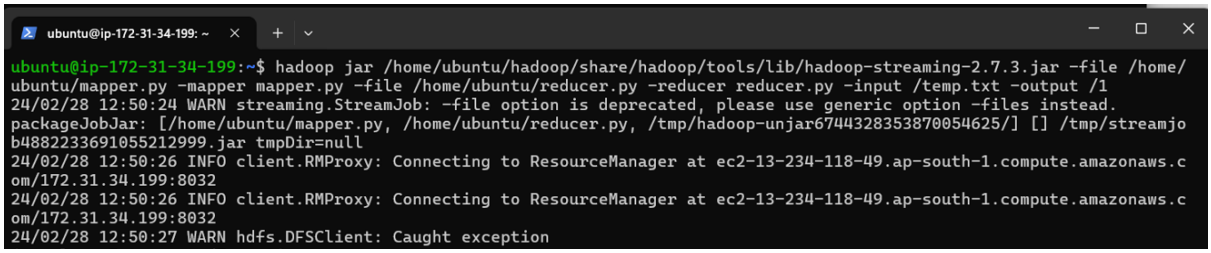
```
#!/usr/bin/python3 -O

import sys

# Loop through each line in the input
for line in sys.stdin:
    # Remove leading and trailing whitespace
    line = line.strip()
    # Split the line into words
    words = line.split()
    # Emit key-value pairs of word and count of 1
    for word in words:
        print(word,"\t",1)
```

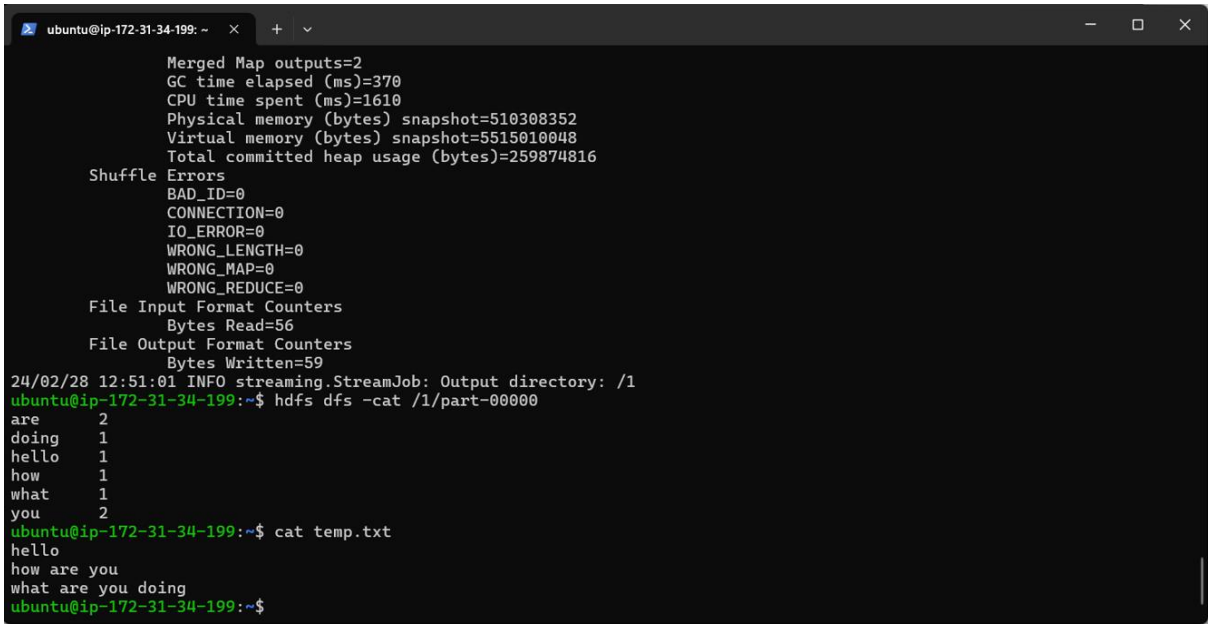
Reducer:

```
#!/usr/bin/python3 -O

import sys

# Initialize variables to keep track of current word and its count
current_word = None
current_count = 0

# Loop through each line in the input
```

```python
for line in sys.stdin:
    # Split the line into word and count, separated by tab
    word, count = line.strip().split('\t', 1)

    # Convert count to integer
    count = int(count)

    # If the word is the same as the current word, increment its count
    if word == current_word:
        current_count += count
    else:
        # If the word is different, print the current word and its count
        if current_word:
            print(current_word,"\t",current_count)
        # Update current word and its count
        current_word = word
        current_count = count

# Print the last word and its count
if current_word:
    print(current_word,"\t",current_count)
```






ubuntu@ip-172-31-34-199: ~
Merged Map outputs=2
GC time elapsed (ms)=370
CPU time spent (ms)=1610
Physical memory (bytes) snapshot=510308352
Virtual memory (bytes) snapshot=5515010048
Total committed heap usage (bytes)=259874816
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=56
File Output Format Counters
Bytes Written=59
24/02/28 12:51:01 INFO streaming.StreamJob: Output directory: /1
ubuntu@ip-172-31-34-199:~$ hdfs dfs -cat /1/part-00000
are 2
doing 1
hello 1
how 1
what 1
you 2
ubuntu@ip-172-31-34-199:~$ cat temp.txt
hello
how are you
what are you doing
ubuntu@ip-172-31-34-199:~$