**Prince Titiya**

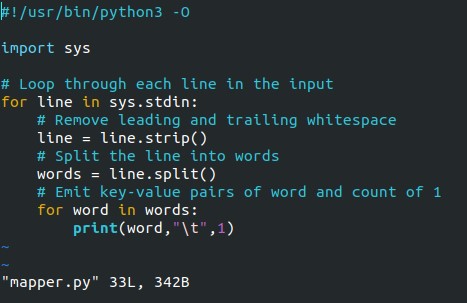
**202318010**

**Big Data Assignment-3**

# Mapper.py

**Shebang**

**Line:**



* Specifies Python 3 interpreter.

**Importing sys Module:**

* sys module for system-specific functions.
* Provides access to system parameters.

**Loop through Input Lines:**

* Iterates over each line of input.
* Reads from standard input.

**Stripping Whitespace:**

* Removes leading and trailing whitespace.
* Ensures clean input. **Splitting Lines into Words:**
* Splits lines into words.
* Based on whitespace.

**Emitting Key-Value Pairs:**

* Prints word and 1.
* Separated by a tab.

# Reducer.py

**Shebang**

**Line:**

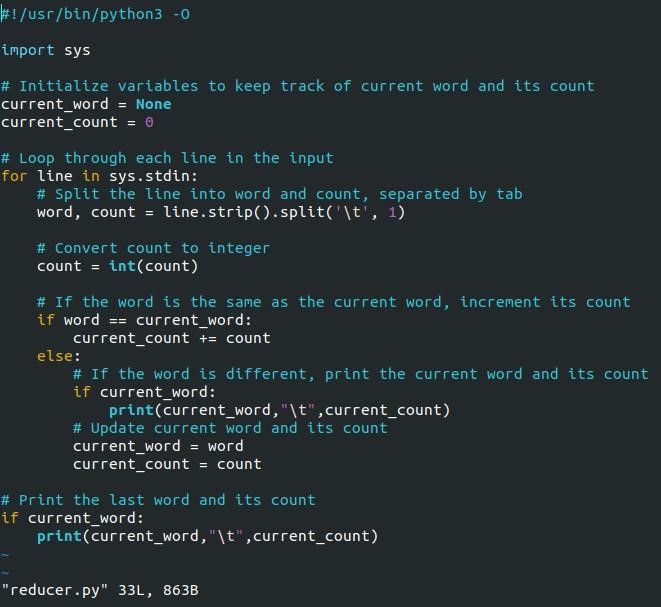
●

Specifies

Python

3

interpreter.



**Importing sys Module:**

* sys module for system-specific functions.
* Provides access to system parameters.

**Initialization:**

* Initialize variables for word and count.
* current\_word and current\_count.

**Loop through Input Lines:**

* Iterates over each line of input.
* Reads from standard input.

**Splitting Lines into Word and Count:**

* Splits each line into word and count.
* Separated by tab, limiting to one split.

**Converting Count to Integer:**

* Converts count from string to integer.
* Ensures numerical operations.

**Incrementing Word Count:**

* Increments count if word is the same.
* Accumulates count for the same word.

**Printing Word Count:**

* Prints current word and count.
* Separated by tab.

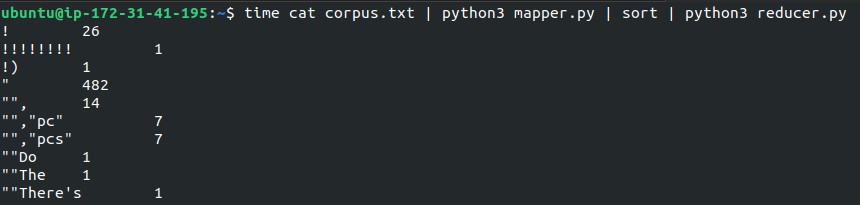
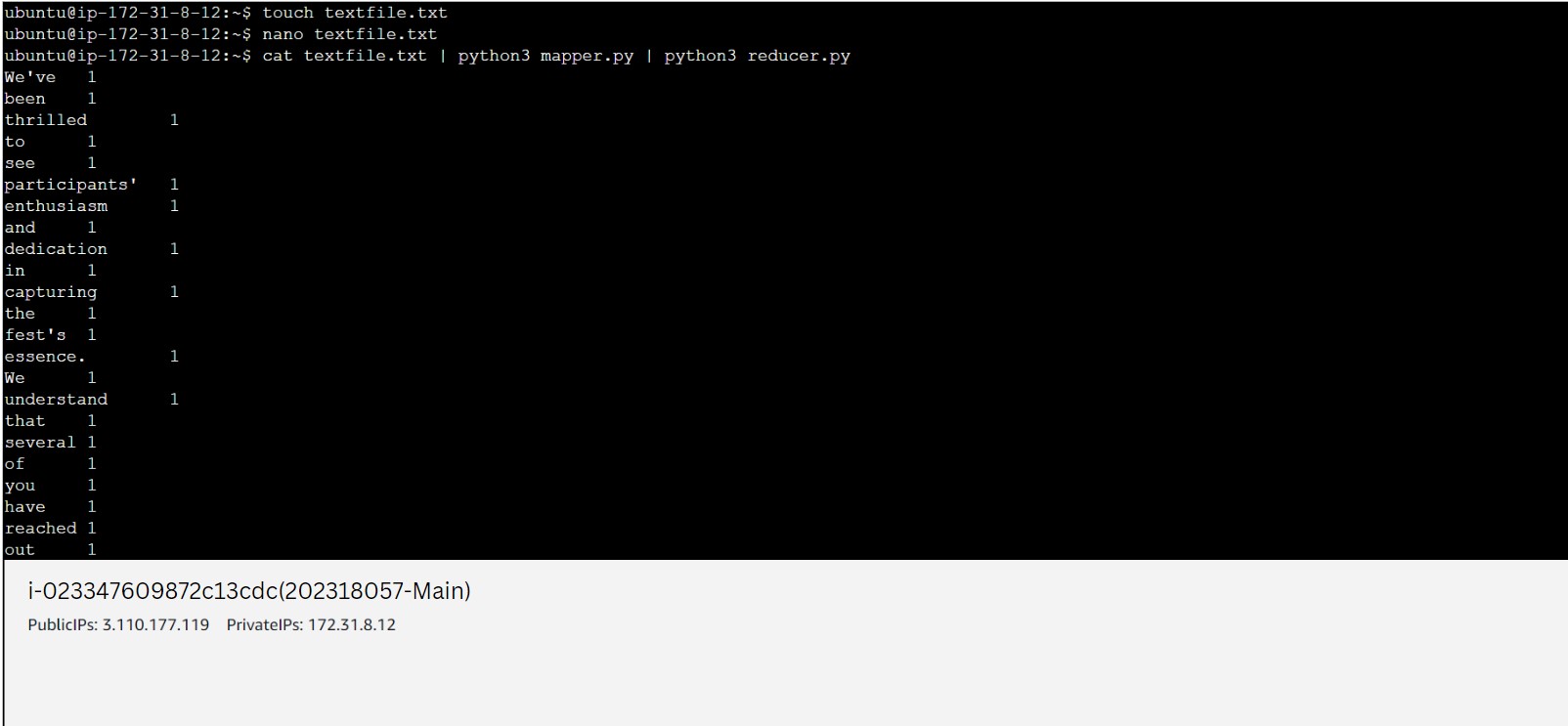
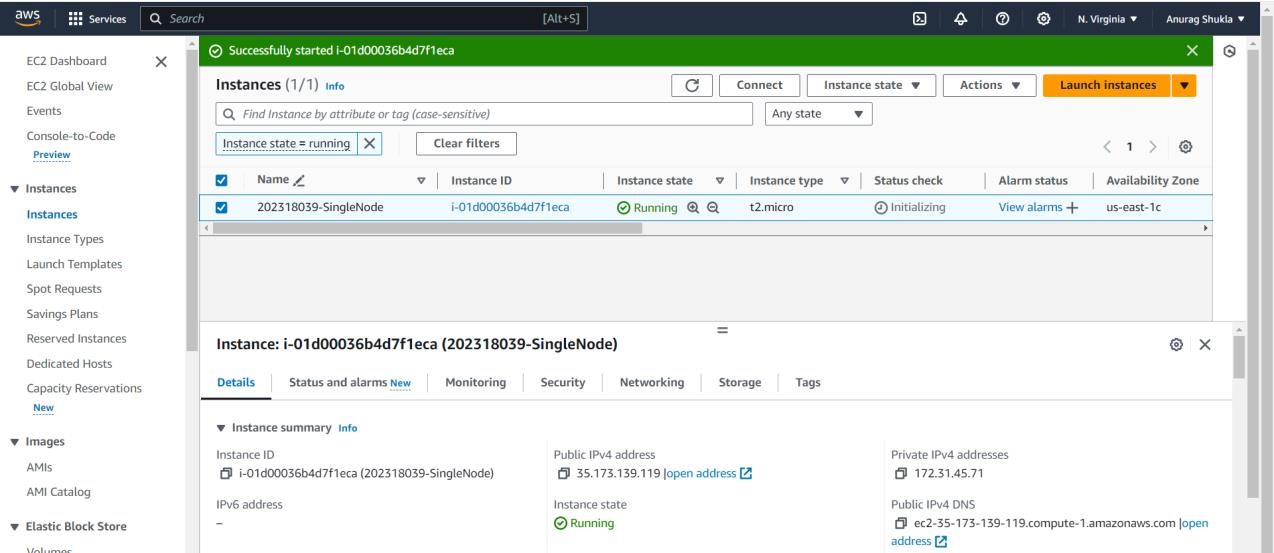
**Updating Current Word and Count:**

* Updates current word and its count.
* Prepares for the next word.

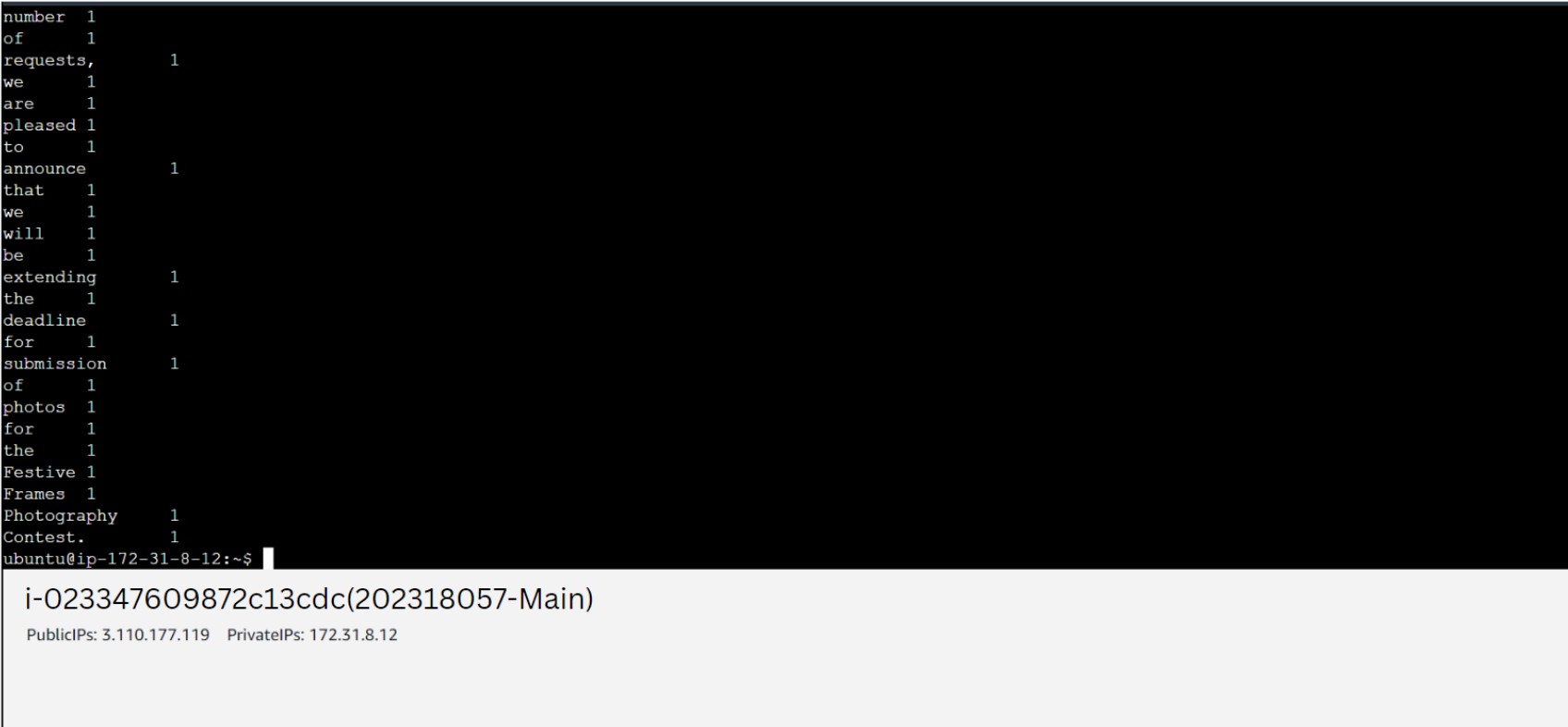
**Printing Last Word and Count:**

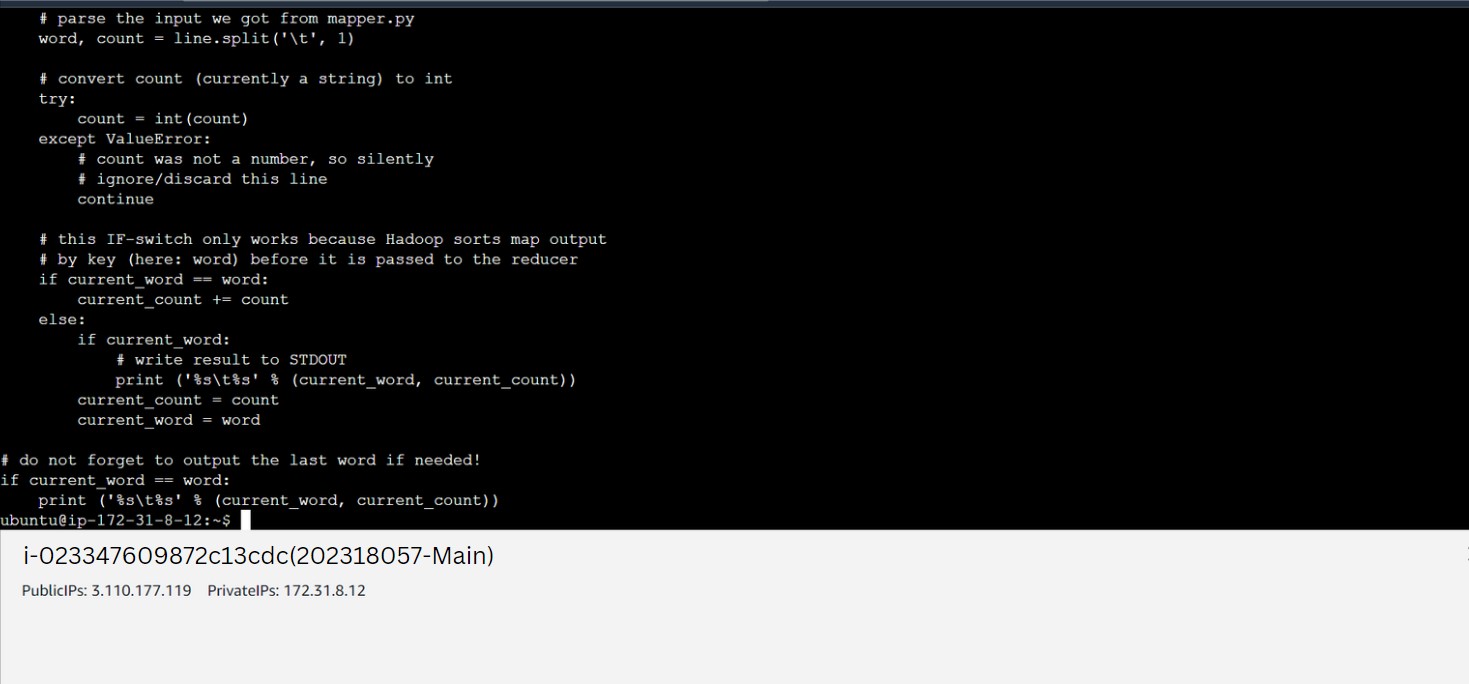
* Prints the last word and its count.
* Ensures all counts are accounted for.

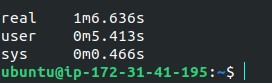
# SingleNode , used txt file corpus.txt ~ 90 mb



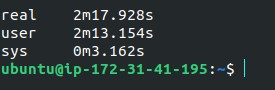
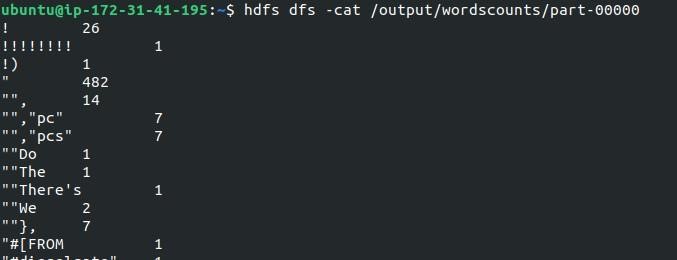
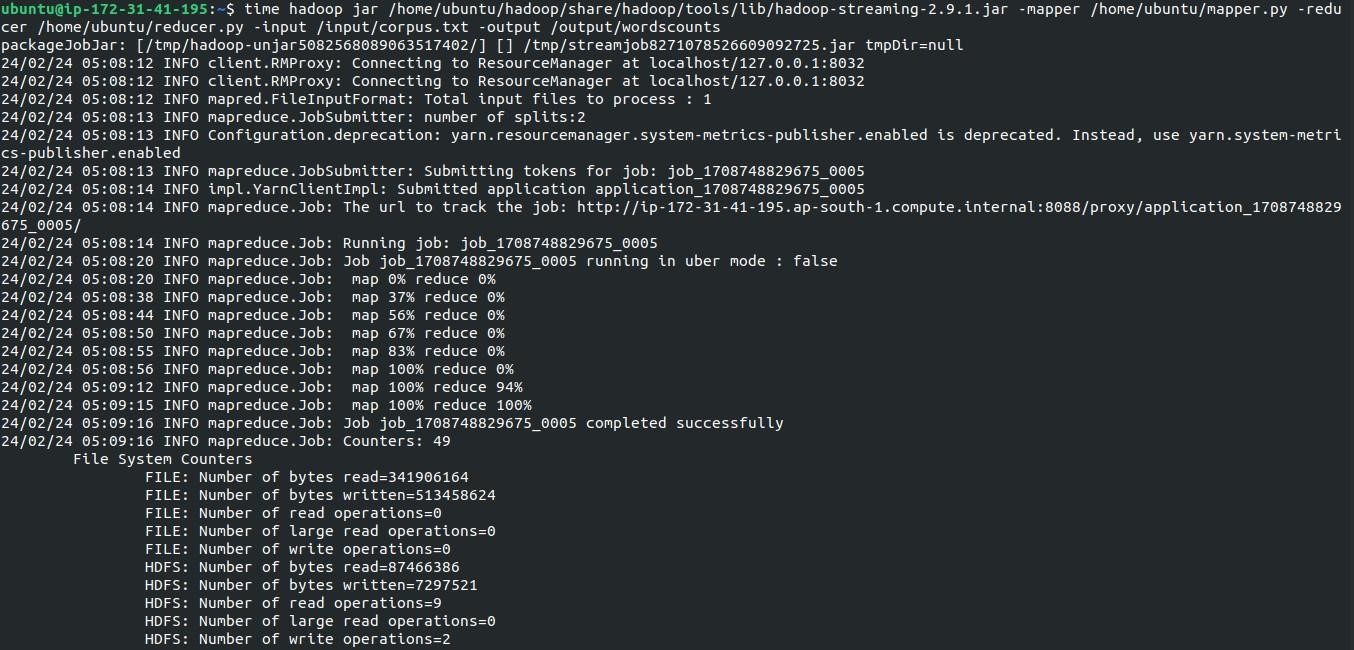
**Without Hadoop**



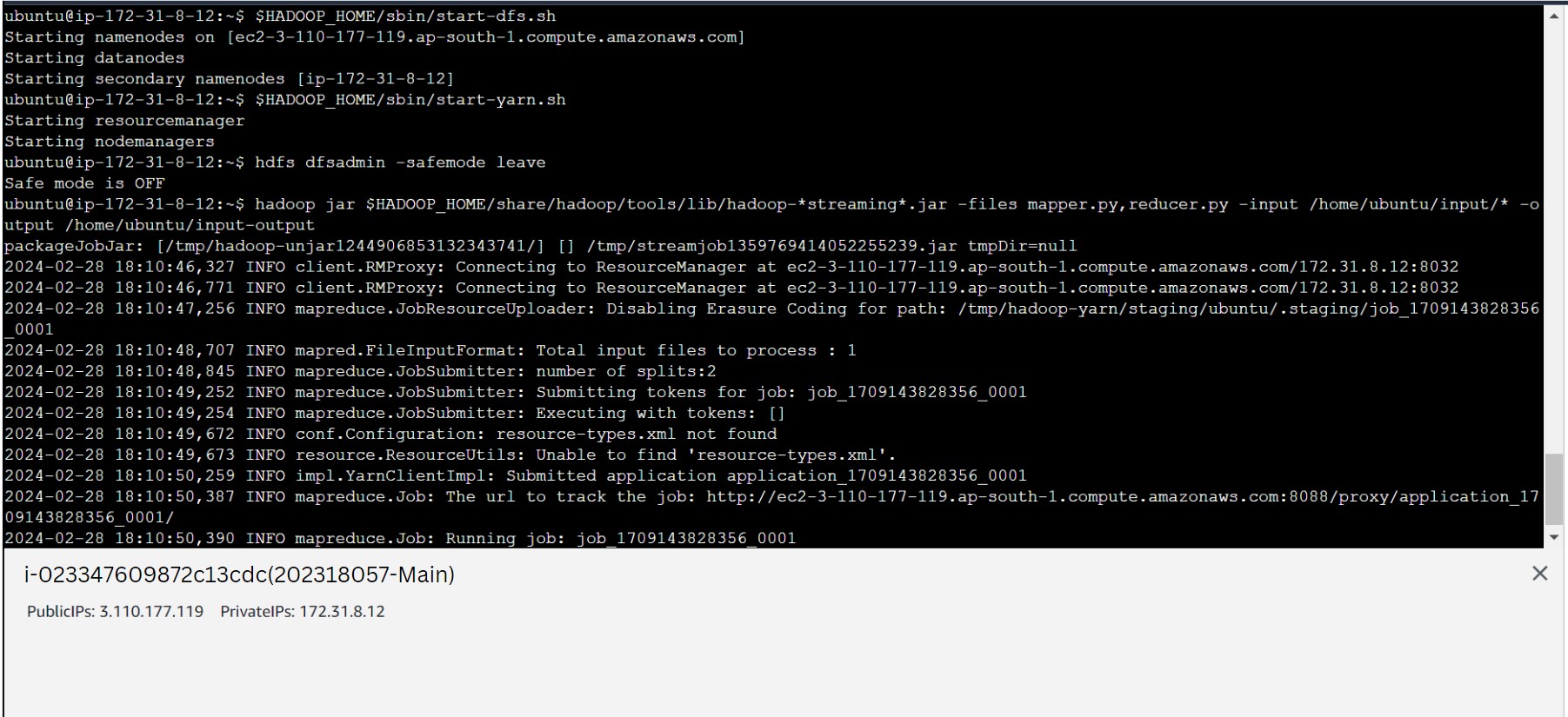
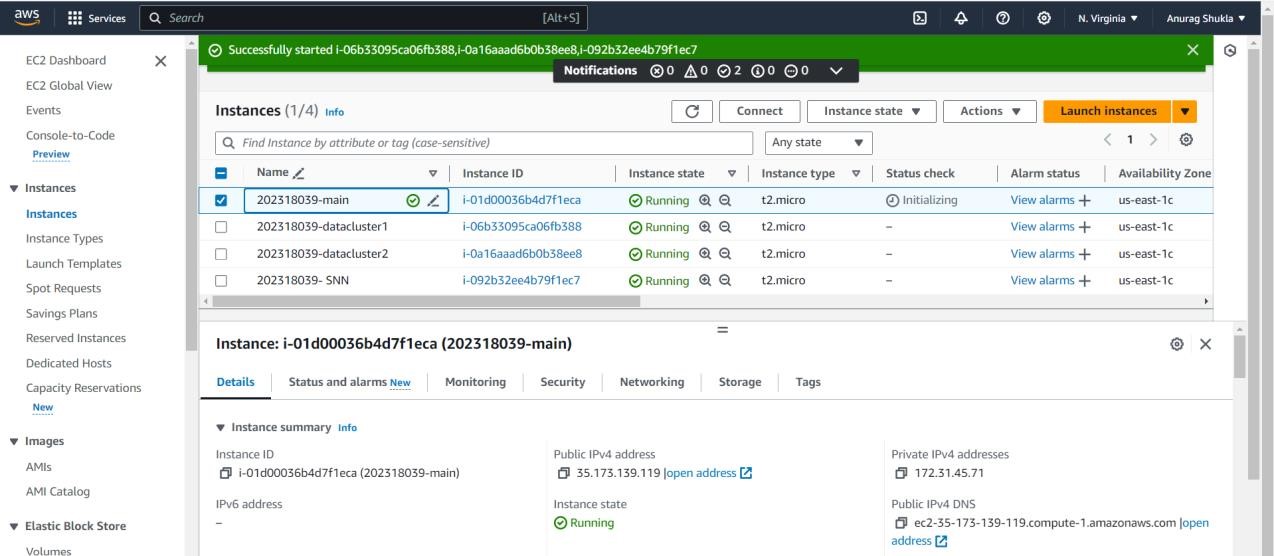




**With Hadoop**



# MultiNode , used txt file corpus.txt ~ 90 mb



**With**

**Hadoop**

