

# ASSIGNMENT-4

## MAP REDUCE

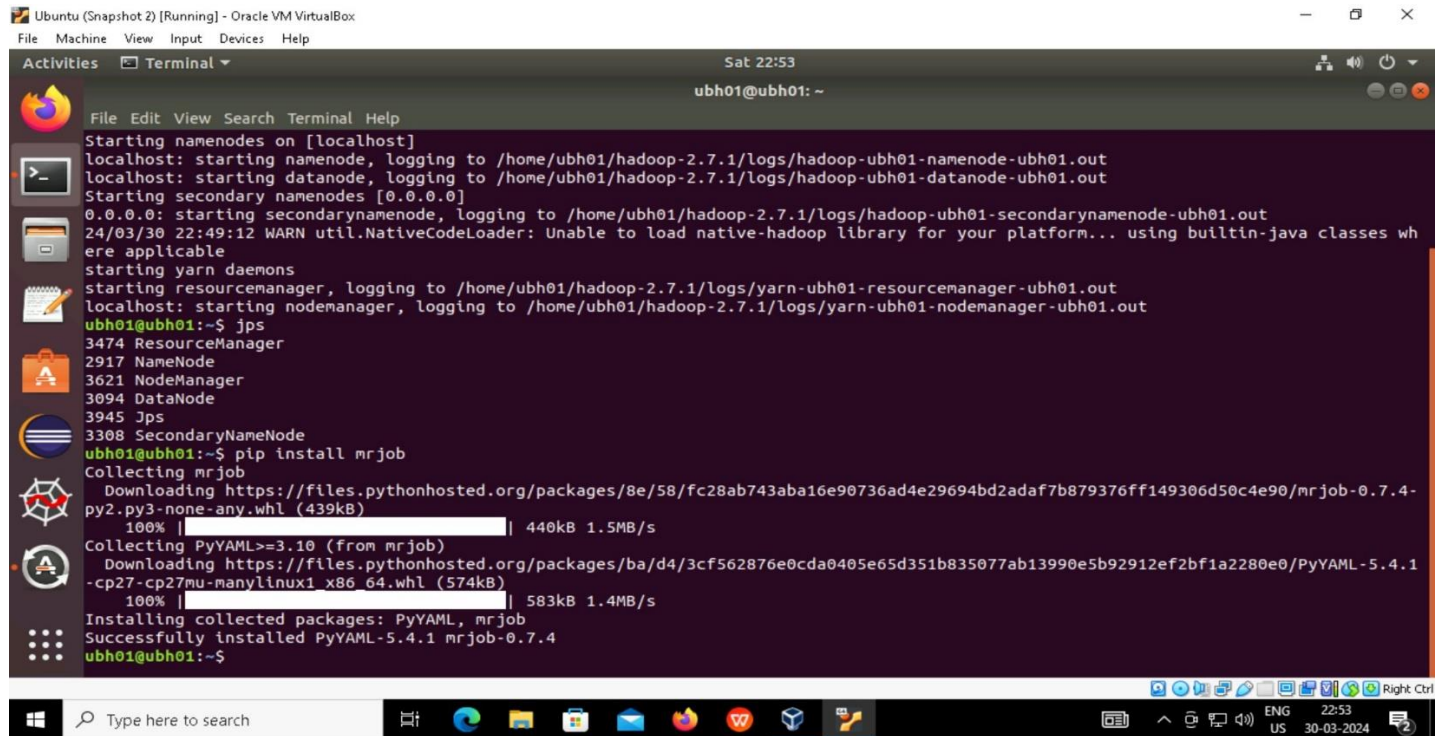
**Name:** Induja D

**Empid:** 2319849

**Cohort Code:** CSDAIA24AZ003

## Assignment 4 MapReduce:

Create one map reduce job to get count of records in given statement:  
**Dear, Bear, River, Car, Car, River, Deer, Car and Bear**



Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox

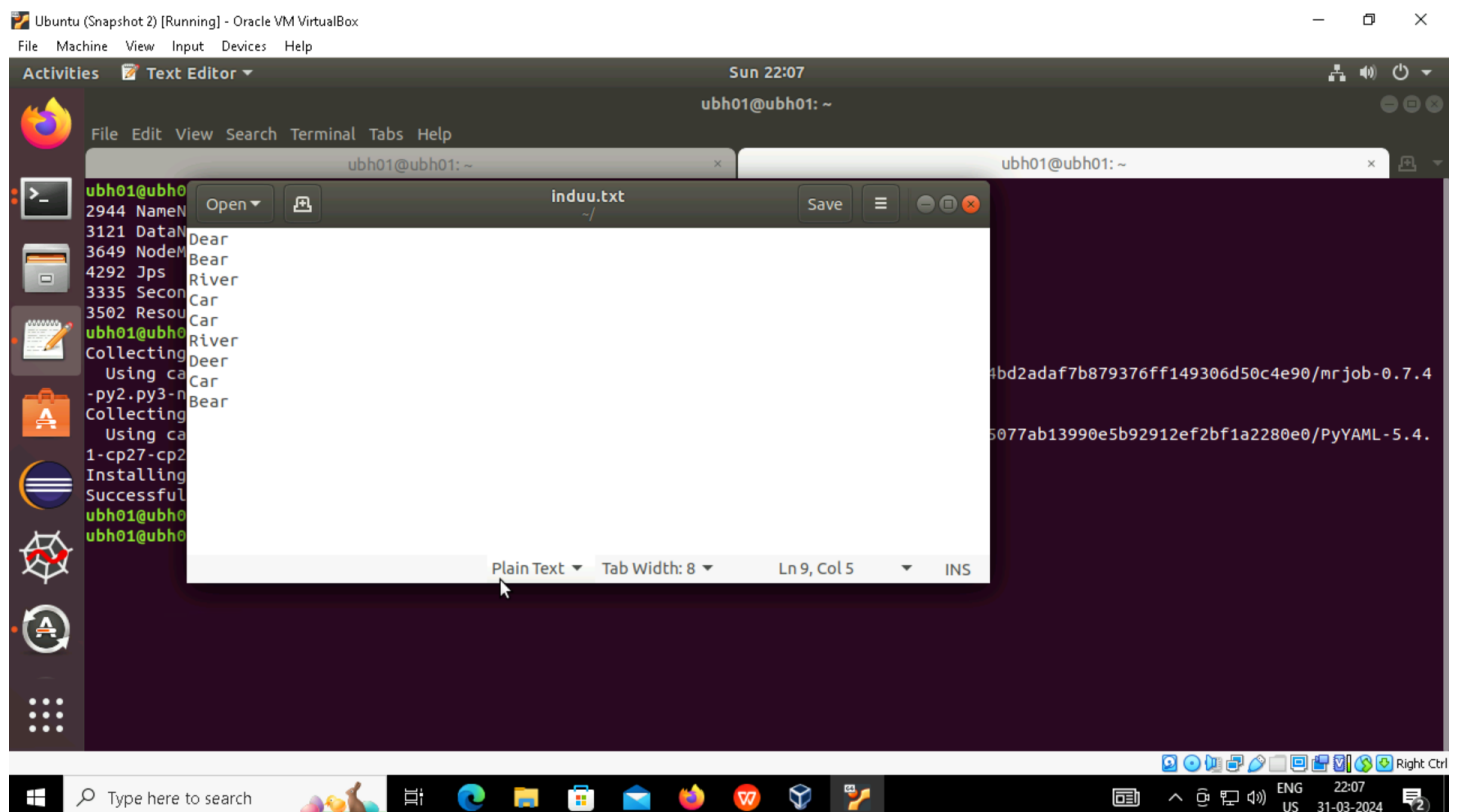
File Machine View Input Devices Help

Activities Terminal

Sat 22:53

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-namenode-ubh01.out
localhost: starting datanode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-datanode-ubh01.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-secondarynamenode-ubh01.out
24/03/30 22:49:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-resourcemanager-ubh01.out
localhost: starting nodemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-nodemanager-ubh01.out
ubh01@ubh01:~$ jps
3474 ResourceManager
2917 NameNode
3621 NodeManager
3094 DataNode
3945 Jps
3308 SecondaryNameNode
ubh01@ubh01:~$ pip install mrjob
Collecting mrjob
  Downloading https://files.pythonhosted.org/packages/8e/58/fc28ab743aba16e90736ad4e29694bd2adaf7b879376ff149306d50c4e90/mrjob-0.7.4-py2.py3-none-any.whl (439kB)
    100% |#####| 440kB 1.5MB/s
Collecting PyYAML>=3.10 (from mrjob)
  Downloading https://files.pythonhosted.org/packages/ba/d4/3cf562876e0cda0405e65d351b835077ab13990e5b92912ef2bf1a2280e0/PyYAML-5.4.1-cp27-cp27mu-manylinux1_x86_64.whl (574kB)
    100% |#####| 583kB 1.4MB/s
Installing collected packages: PyYAML, mrjob
Successfully installed PyYAML-5.4.1 mrjob-0.7.4
ubh01@ubh01:~$
```



Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Text Editor

Sun 22:07

ubh01@ubh01: ~

ubh01@ubh01: ~

induu.txt

```
2944 NameN
3121 DataN
3649 NodeM
4292 Jps
3335 Secon
3502 Resou
ubh01@ubh0
Collecting
Using ca
-py2.py3-n
Collecting
Using ca
1-cp27-cp2
Installing
Successful
ubh01@ubh0
ubh01@ubh0
```

Dear  
Bear  
River  
Car  
Car  
River  
Deer  
Car  
Bear

Plain Text Tab Width: 8 Ln 9, Col 5 INS

Sun 22:12

ubh01@ubh01: ~

File Edit View Search Terminal Tabs Help

ubh01@ubh01: ~

Open ▾

induu.txt

\*count\_indu.py

```

from mrjob.job import MRJob
from mrjob.step import MRStep

class WordBreakdown(MRJob):
    def steps(self):
        return [
            MRStep(mapper=self.mapper_get_word,
                  reducer=self.reducer_count_word)
        ]

    def mapper_get_word(self, _, line):
        word = line
        yield word, 1

    def reducer_count_word(self, key, values):
        yield key, sum(values)

if __name__ == '__main__':
    WordBreakdown.run()

```

Python ▾ Tab Width: 8 ▾ Ln 13, Col 21 ▾ INS

Sun 22:13

ubh01@ubh01: ~

File Edit View Search Terminal Tabs Help

ubh01@ubh01: ~

```

4292 Jps
3335 SecondaryNameNode
3502 ResourceManager
ubh01@ubh01:~$ pip install mrjob
Collecting mrjob
  Using cached https://files.pythonhosted.org/packages/8e/58/fc28ab743aba16e90736ad4e29694bd2ada7b7879376ff149306d50c4e90/mrjob-0.7.4-py2.py3-none-any.whl
Collecting PyYAML>=3.10 (from mrjob)
  Using cached https://files.pythonhosted.org/packages/ba/d4/3cf562876e0cda0405e65d351b835077ab13990e5b92912ef2bf1a2280e0/PyYAML-5.4.1-cp27-cp27mu-manylinux1_x86_64.whl
Installing collected packages: PyYAML, mrjob
Successfully installed PyYAML-5.4.1 mrjob-0.7.4
ubh01@ubh01:~$ gedit induu.txt
ubh01@ubh01:~$ gedit count_indu.py
ubh01@ubh01:~$ python count_indu.py induu.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/count_indu.ubh01.20240331.164305.046293
Running step 1 of 1...
job output is in /tmp/count_indu.ubh01.20240331.164305.046293/output
Streaming final output from /tmp/count_indu.ubh01.20240331.164305.046293/output...
"Car" 3
"River" 2
"Dear" 1
"Deer" 1
"Bear" 2
Removing temp directory /tmp/count_indu.ubh01.20240331.164305.046293...
ubh01@ubh01:~$

```

