



ASSIGNMENT-4

(MAP REDUCE)

Employee ID : 2320074

CSDAIA24AZ003

```
Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Sun 10:37
ubh01@ubh01: ~

File Edit View Search Terminal Help
ubh01@ubh01:~$ jps
2597 Jps
ubh01@ubh01:~$ ./hadoop-2.7.1/sbin/startall.sh
bash: ./hadoop-2.7.1/sbin/startall.sh: No such file or directory
ubh01@ubh01:~$ ./hadoop-2.7.1/sbin/start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
24/03/31 10:37:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
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
24/03/31 10:37:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere 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:~$
```

```
Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Sun 10:40
ubh01@ubh01: ~

File Edit View Search Terminal Help
24/03/31 10:37:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
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
24/03/31 10:37:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere 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:~$ echo "Dear Bear River Car Car River Deer Car Bear" >>dharani.txt
ubh01@ubh01:~$ ls
apache-hive-2.3.2-bin          hadoop-2.7.1                  Public
apache-hive-2.3.2-bin.tar.gz  hadoop-2.7.1.tar.gz          pyspark
'Bigdata Case-Studies_Data Files' hbase                         scala-SDK-4.7.0-vfinal-2.12-linux.gtk.x86_64
cs                             hbase-1.1.2                  scala-SDK-4.7.0-vfinal-2.12-linux.gtk.x86_64.tar.gz
derby.log                     kafka_2.12-2.8.0              spark-3.0.3-bin-hadoop2.7
Desktop                        kafka_2.12-2.8.0.tgz          spark-3.0.3-bin-hadoop2.7.tgz
dfsdata                       'LAB_ATHDW331127(Pyspark)-Data Files' spark-warehouse
dharani.txt                   'LAB_ATHDW331340(Apache Spark)-Data Files' sqoop-1.4.7.bin__hadoop-2.6.0
Downloads                     metastore_db                  sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
eclipse                       Music                         Templates
eclipse-installer             nifi-1.15.3-bin              TestFile.txt
eclipse-workspace             nifi-1.15.3-bin.tar.gz       tmpdata
examples.desktop              Pictures                       Videos
                               workspace
```

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

File Machine View Input Devices Help

Activities Terminal

Sun 10:40

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
24/03/31 10:37:47 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:~$ echo "Dear Bear River Car Car River Deer Car Bear" >>dharanti.txt
ubh01@ubh01:~$ ls
apache-hive-2.3.2-bin          hadoop-2.7.1                  Public
apache-hive-2.3.2-bin.tar.gz  hadoop-2.7.1.tar.gz           pyspark
'Bigdata Case-Studies_Data Files' HBASE                         scala-SDK-4.7.0-vfinal-2.12-linux.gtk.x86_64
cs                             hbase-1.1.2                  scala-SDK-4.7.0-vfinal-2.12-linux.gtk.x86_64.tar.gz
derby.log                     hbase-1.1.2-bin.tar.gz       spark-3.0.3-bin-hadoop2.7
Desktop                       kafka_2.12-2.8.0              spark-3.0.3-bin-hadoop2.7.tgz
dfsdata                       kafka_2.12-2.8.0.tgz          spark-warehouse
dharanti.txt                  'LAB_ATHDW331127(Pyspark)-Data Files' sqoop-1.4.7.bin__hadoop-2.6.0
Documents                     'LAB_ATHDW331340(Apache Spark)-Data Files' sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
Downloads                     metastore_db                  Templates
eclipse                       Music                         TestFile.txt
eclipse-installer             nifi-1.15.3-bin              tmpdata
eclipse-workspace             nifi-1.15.3-bin.tar.gz       Videos
examples_desktop              Pictures                       workspace
ubh01@ubh01:~$ cat dharanti.txt
Dear Bear River Car Car River Deer Car Bear
ubh01@ubh01:~$
```

Type here to search

Right Ctrl

ENG 10:40 31-03-2024

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

File Machine View Input Devices Help

Activities Text Editor

Sun 10:47

pythonfile.py

```
Open Save
from mrjob.job import MRJob
from mrjob.step import MRStep
class MRWordCount(MRJob):
    def mapper(self,_,line):
        for word in line.split():
            yield word.lower(), 1
    def reducer(self,word,counts):
        yield word,sum(counts)
    def steps(self):
        return [
            MRStep(mapper=self.mapper,reducer=self.reducer)
        ]
if __name__=='__main__':
    MRWordCount.run()
```

Python Tab Width: 8 Ln 14, Col 26 INS

Type here to search

Right Ctrl

ENG 10:47 31-03-2024

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

File Machine View Input Devices Help

Activities Terminal ▾ Sun 10:48

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
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 1.9MB/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:~$
```

Windows taskbar: Type here to search, 10:48, 31-03-2024, ENG US

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

File Machine View Input Devices Help

Activities Terminal ▾ Sun 10:49

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
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 1.9MB/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:~$ python pythonfile.py dharani.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/pythonfile.ubh01.20240331.051916.997506
Running step 1 of 1...
job output is in /tmp/pythonfile.ubh01.20240331.051916.997506/output
Streaming final output from /tmp/pythonfile.ubh01.20240331.051916.997506/output...
"car" 3
"river" 2
"deer" 1
"deer" 1
"bear" 2
Removing temp directory /tmp/pythonfile.ubh01.20240331.051916.997506...
ubh01@ubh01:~$
```

Windows taskbar: Type here to search, 10:49, 31-03-2024, ENG US

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

File Machine View Input Devices Help

Activities Terminal ▾

Sun 10:51

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
ubh01@ubh01:~$ hdfs dfs -put pythonfile.py /
24/03/31 10:50:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hdfs dfs -put dharani.txt /
24/03/31 10:51:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$
```

Type here to search

ENG 10:51 US 31-03-2024

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

File Machine View Input Devices Help

Activities Terminal ▾

Sun 10:51

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
ubh01@ubh01:~$ hdfs dfs -put pythonfile.py /
24/03/31 10:50:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hdfs dfs -put dharani.txt /
24/03/31 10:51:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hdfs dfs -ls /
24/03/31 10:51:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
-rw-r--r-- 1 ubh01 supergroup 44 2024-03-31 10:51 /dharani.txt
-rw-r--r-- 1 ubh01 supergroup 349 2024-03-31 10:50 /pythonfile.py
drwx-wx-wx - ubh01 supergroup 0 2022-01-20 18:25 /tmp
drwxr-xr-x - ubh01 supergroup 0 2022-01-20 17:54 /user
ubh01@ubh01:~$
```

Type here to search

ENG 10:51 US 31-03-2024


```
Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Sun 10:53
ubh01@ubh01: ~

File Edit View Search Terminal Help
ubh01@ubh01:~$ python pythonfile.py -r hadoop hdfs:///dharan1.txt
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in /home/ubh01/hadoop-2.7.1/bin...
Found hadoop binary: /home/ubh01/hadoop-2.7.1/bin/hadoop
Using Hadoop version 2.7.1
Looking for Hadoop streaming jar in /home/ubh01/hadoop-2.7.1...
Found Hadoop streaming jar: /home/ubh01/hadoop-2.7.1/share/hadoop/tools/lib/hadoop-streaming-2.7.1.jar
Creating temp directory /tmp/pythonfile.ubh01.20240331.052322.132301
uploading working dir files to hdfs:///user/ubh01/tmp/mrjob/pythonfile.ubh01.20240331.052322.132301/files/wd...
```

```

Type here to search
10:53
ENG
US
31-03-2024

Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Sun 10:54
ubh01@ubh01: ~

File Edit View Search Terminal Help
Merged Map outputs=2
Physical memory (bytes) snapshot=709808128
Reduce input groups=5
Reduce input records=9
Reduce output records=5
Reduce shuffle bytes=110
Shuffled Maps =2
Spilled Records=18
Total committed heap usage (bytes)=547356672
Virtual memory (bytes) snapshot=5753810944
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
job output is in hdfs:///user/ubh01/tmp/mrjob/pythonfile.ubh01.20240331.052322.132301/output
Streaming final output from hdfs:///user/ubh01/tmp/mrjob/pythonfile.ubh01.20240331.052322.132301/output...
"bear" 2
"car" 3
"deer" 1
"deer" 1
"river" 2
STDERR: 24/03/31 10:54:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Removing HDFS temp directory hdfs:///user/ubh01/tmp/mrjob/pythonfile.ubh01.20240331.052322.132301...
Removing temp directory /tmp/pythonfile.ubh01.20240331.052322.132301...
ubh01@ubh01:~$
```