

Ex.No – 10

Date – 27/08/2024

DEMONSTRATE THE MAP REDUCE PROGRAMMING MODEL BY COUNTING THE NUMBER OF WORDS IN A FILE

AIM:

To demonstrate the MAP REDUCE programming model for counting the number of words in a file.

PROCEDURE:

Step 1 - Open Terminal

Step 2 - Start dfs and mapreduce services

```
$ cd /opt/homebrew/Cellar/hadoop/3.4.0/libexec/etc/hadoop
```

```
$ Hadoop namenode -format
```

```
$ start-all.sh
```

```
$ jps
```

Step 3 - Check Hadoop through web UI

// Go to browser type <http://localhost:9087> – All Applications Hadoop Cluster

Step 4 - Make a input.txt file and 2 python file namely mapper.py and reducer.py

Step 5 - Create a directory in hadoop

```
Hadoop fs -mkdir /user
```

Step 6 - Upload the input text file in the user directory.

```
Hadoop fs -put path_to_input /user
```

Step 7 - Now execute the mapper and the reducer python file using following commands.

```
hadoop jar
```

```
/opt/homebrew/Cellar/hadoop/3.4.0/libexec/share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar \
```

```
-files
```

```
/Users/madhumithak/Documents/DA/Lab/Ex3/mapper.py,/Users/madhumithak/Documents/DA/Lab/Ex3/reducer.py \
```

```
-input /user/input.txt \
```

```
-output /user/output \
```

```
-mapper "python3 mapper.py" \
```

```
-reducer "python3 reducer.py"
```

Step 8 - the output is seen using the following commands

```
hadoop fs -cat /user/output/part-00000
```

PROGRAM -

input.txt -

hello world

Madhumitha K

Computer Science Engineering

Rajalakshmi Engineering College

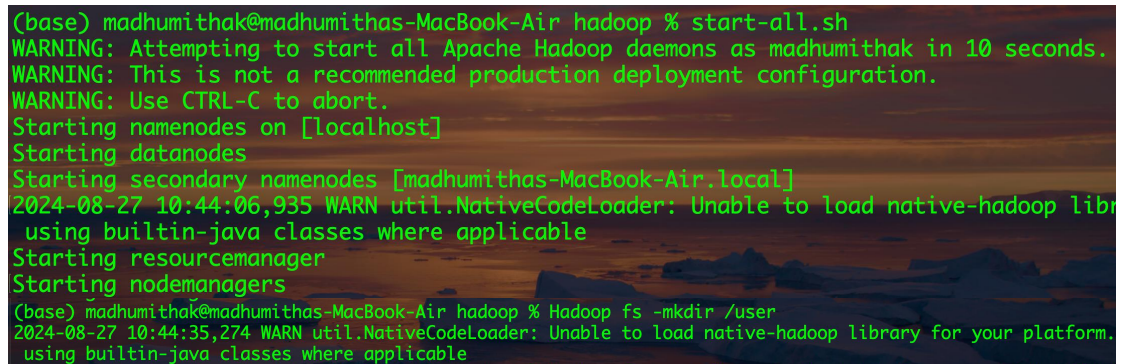
mapper.py -

```
#!/usr/bin/python3
import sys
for line in sys.stdin:
    line=line.strip()
    words=line.split()
    for word in words:
        print ('%s\t%s'%(word, 1))
```

reducer.py -

```
#!/usr/bin/python3
import sys
prevw=None
prevc=0
for line in sys.stdin:
    line = line.strip()
    word, count= line.split('\t')
    count=int(count)
    if prevw == word:
        prevc += count
    else:
        if prevw:
            print ('%s\t%s' % (prevw, prevc))
            prevc = count
            prevw = word
if prevw == word:
    print ('%s\t%s' % (prevw, prevc) )
```

OUTPUT :

A terminal window screenshot showing the execution of 'start-all.sh' in a Hadoop environment. The background of the terminal has a dark, abstract image. The text is green on a black background. It shows warnings about starting daemons and a production configuration, followed by the startup of namenodes, datanodes, and nodemanagers. A final warning about native-hadoop library is also present.

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as madhumithak in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [madhumithas-MacBook-Air.local]
2024-08-27 10:44:06,935 WARN util.NativeCodeLoader: Unable to load native-hadoop library
using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers
(base) madhumithak@madhumithas-MacBook-Air hadoop % Hadoop fs -mkdir /user
2024-08-27 10:44:35,274 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
using builtin-java classes where applicable
```

Browse Directory

/

Go!

Show

25

entries

Search:

<input type="checkbox"/>	<div></div> Permission	<div></div> Owner	<div></div> Group	<div></div> Size	<div></div> Last Modified	<div></div> Replication	<div></div> Block Size	<div></div> Name	<div></div>
<input type="checkbox"/>	drwxr-xr-x	madhumithak	supergroup	0 B	Aug 27 10:45	<u>0</u>	0 B	user	<div></div>

Showing 1 to 1 of 1 entries

Previous

1

Next

Hadoop, 2024.

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % Hadoop fs -put /Users/madhumithak/Documents/DA/Lab/Ex2/input.txt /user
2024-08-27 10:45:40,000 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
(base) madhumithak@madhumithas-MacBook-Air hadoop %
```

Browse Directory

/user

Go!

Show

25

entries

Search:

<input type="checkbox"/>	<div></div> Permission	<div></div> Owner	<div></div> Group	<div></div> Size	<div></div> Last Modified	<div></div> Replication	<div></div> Block Size	<div></div> Name	<div></div>
<input type="checkbox"/>	-rw-r--r--	madhumithak	supergroup	87 B	Aug 27 10:45	<u>1</u>	128 MB	input.txt	<div></div>

Showing 1 to 1 of 1 entries

Previous

1

Next

Hadoop, 2024.

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % hadoop jar /opt/homebrew/Cellar/hadoop/3.4.0/libexec/share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar \
-files /Users/madhumithak/Documents/DA/Lab/Ex2/mapper.py,/Users/madhumithak/Documents/DA/Lab/Ex2/reducer.py \
-input /user/input.txt \
-output /user/output \
-mapper "python3 mapper.py" \
-reducer "python3 reducer.py"
2024-08-27 10:51:58,928 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
2024-08-27 10:51:59,631 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-08-27 10:51:59,730 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-08-27 10:51:59,730 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-08-27 10:51:59,798 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-08-27 10:52:00,256 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-27 10:52:00,438 INFO mapreduce.JobSubmitter: number of splits:1
2024-08-27 10:52:00,615 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local2075277960_0001
2024-08-27 10:52:00,615 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-27 10:52:00,866 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-madhumithak/mapred/local/job_local2075277960_0001_e677e186-83d4-4e1e-8a7d-c455aca6ba1d/mapper.py <- /opt/homebrew/Cellar/hadoop/3.4.0/libexec/etc/hadoop/mapper.py
2024-08-27 10:52:00,876 INFO mapred.LocalDistributedCacheManager: Localized file:/Users/madhumithak/Documents/DA/Lab/Ex2/mapper.py as file:/tmp/hadoop-madhumithak/mapred/local/job_local2075277960_0001_e677e186-83d4-4e1e-8a7d-c455aca6ba1d/mapper.py
2024-08-27 10:52:00,905 INFO mapred.LocalDistributedCacheManager: Creating symlink: /tmp/hadoop-madhumithak/mapred/local/job_local2075277960_0001_e677e186-83d4-4e1e-8a7d-c455aca6ba1d/reducer.py <- /opt/homebrew/Cellar/hadoop/3.4.0/libexec/etc/hadoop/reducer.py
```

Browse Directory

Show
25
entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	madhumithak	supergroup	87 B	Aug 27 10:45	1	128 MB	input.txt	
<input type="checkbox"/>	drwxr-xr-x	madhumithak	supergroup	0 B	Aug 27 10:52	0	0 B	output	

Showing 1 to 2 of 2 entries

Previous
1
Next

Hadoop, 2024.

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % hadoop fs -cat /user/output/part-00000
2024-08-27 10:53:14,395 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
using builtin-java classes where applicable
College 1
Computer      1
Engineering   2
K             1
Madhumitha    1
Rajalakshmi   1
Science 1
hello 1
world 1
(base) madhumithak@madhumithas-MacBook-Air hadoop %
```

RESULT:

Thus the map reducer program to count the number of words in a file is executed successfully.