Ex No 2

Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

AIM:

To run a basic Word Count MapReduce program using Hadoop.

PROCEDURE:

Step 1: Start the Hadoop cluster

1. Open Terminal in administrative mode:
    - Open a terminal window.
    - Run Hadoop's startup scripts to start the cluster:

```
 cd /usr/local/Cellar/hadoop/3.4.0/libexec/sbin
```

```
./start-dfs.sh
./start-yarn.sh
```

2. Verify that all nodes are up by running:

```
jps
```

Step 2: Create an input directory in HDFS

Create an HDFS directory where you will place the input file for the MapReduce job. You can name it "input_dir":

```
hadoop fs -mkdir /input_dir
```

Step 3: Copy the input text file to the input directory

Prepare your input file (named `input_file.txt`), or create a sample text file on your local system:

```
echo "Hadoop is a distributed computing framework" >
~/input_file.txt
```

Copy the input file to HDFS:

```
hadoop fs -put ~/input_file.txt /input_dir
```

Step 4: Verify if the file is copied to HDFS
List files in the input directory:

```
hadoop fs -ls /input_dir
```

Check the content of the copied file:

```
hadoop fs -cat /input_dir/input_file.txt
```

Step 5: Run the MapReduce Word Count job

1. Run the MapReduce job:
    - Use the built-in WordCount example that comes with Hadoop.
    - Run the following command, specifying the input directory (`/input_dir`) and an output directory (`/output_dir`):

```
hadoop jar /usr/local/Cellar/hadoop/3.4.0/libexec/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar wordcount /input_dir /output_dir
```

Step 6: Verify the output generated
Check the content of the output directory:

```
hadoop fs -ls /output_dir
```

View the content of the output file:

```
hadoop fs -cat /output_dir/part-r-00000
```

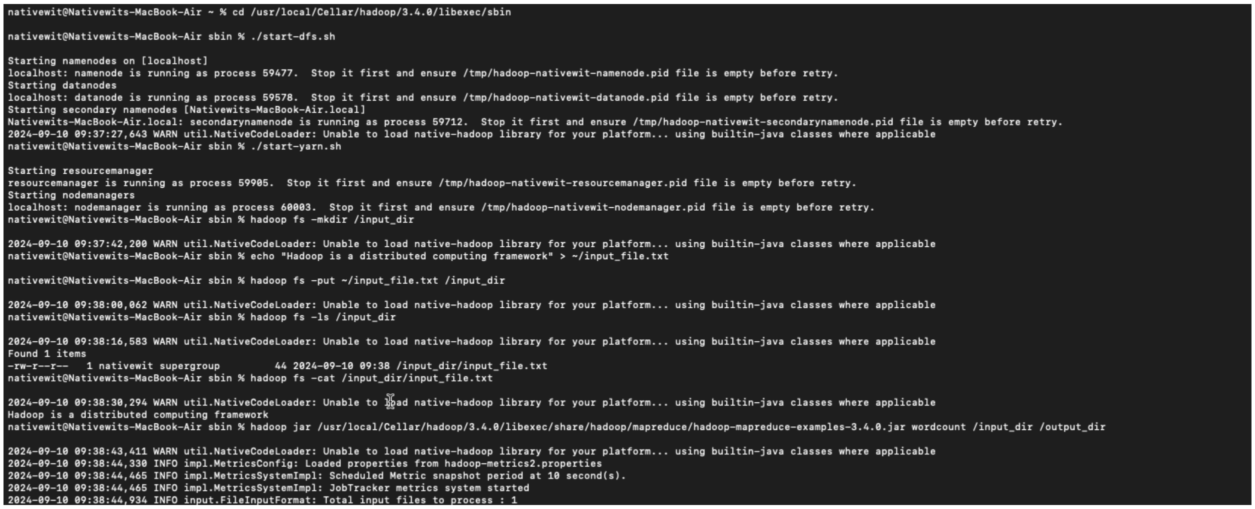
Step 7: Useful Hadoop Commands
To delete a file from HDFS directory:

```
hadoop fs -rm -r /input_dir/input_file.txt
```

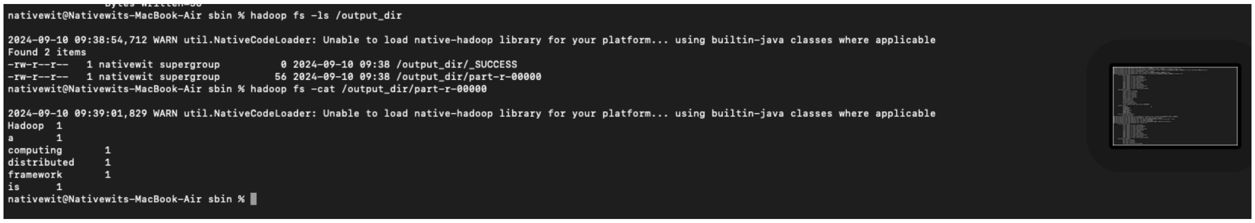
To delete a directory from HDFS directory:

```
hadoop fs -rm -r /input_dir
```

Output:



```
nativewit@Nativewits-MacBook-Air ~ % cd /usr/local/Cellar/hadoop/3.4.0/libexec/sbin

nativewit@Nativewits-MacBook-Air sbin % ./start-dfs.sh

Starting namenodes on [localhost]
localhost: namenode is running as process 59477.  Stop it first and ensure /tmp/hadoop-nativewit-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 59578.  Stop it first and ensure /tmp/hadoop-nativewit-datanode.pid file is empty before retry.
Starting secondary namenodes [Nativewits-MacBook-Air.local]
Nativewits-MacBook-Air.local: secondarynamenode is running as process 59712.  Stop it first and ensure /tmp/hadoop-nativewit-secondarynamenode.pid file is empty before retry.
2024-09-10 09:37:27,643 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % ./start-yarn.sh

Starting resourcemanager
resourcemanager is running as process 59905.  Stop it first and ensure /tmp/hadoop-nativewit-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 60003.  Stop it first and ensure /tmp/hadoop-nativewit-nodemanager.pid file is empty before retry.
nativewit@Nativewits-MacBook-Air sbin % hadoop fs -mkdir /input_dir

2024-09-10 09:37:42,200 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % echo "Hadoop is a distributed computing framework" > ~/input_file.txt

nativewit@Nativewits-MacBook-Air sbin % hadoop fs -put ~/input_file.txt /input_dir

2024-09-10 09:38:00,062 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hadoop fs -ls /input_dir

2024-09-10 09:38:16,583 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 nativewit supergroup         44 2024-09-10 09:38 /input_dir/input_file.txt
nativewit@Nativewits-MacBook-Air sbin % hadoop fs -cat /input_dir/input_file.txt

2024-09-10 09:38:30,294 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hadoop is a distributed computing framework
nativewit@Nativewits-MacBook-Air sbin % hadoop jar /usr/local/Cellar/hadoop/3.4.0/libexec/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar wordcount /input_dir /output_dir

2024-09-10 09:38:43,411 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-09-10 09:38:44,330 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-09-10 09:38:44,465 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-09-10 09:38:44,465 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-09-10 09:38:44,934 INFO input.FileInputFormat: Total input files to process : 1
```



```
nativewit@Nativewits-MacBook-Air sbin % hadoop fs -ls /output_dir

2024-09-10 09:38:54,712 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 nativewit supergroup          0 2024-09-10 09:38 /output_dir/_SUCCESS
-rw-r--r--   1 nativewit supergroup         56 2024-09-10 09:38 /output_dir/part-r-00000
nativewit@Nativewits-MacBook-Air sbin % hadoop fs -cat /output_dir/part-r-00000

2024-09-10 09:39:01,829 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hadoop  1
a       1
computing       1
distributed     1
framework       1
is      1
nativewit@Nativewits-MacBook-Air sbin % 
```

RESULT:

Thus, the program for basic Word Count Map Reduce has been executed successfully.