

## Practical No: 7

Date: 17/04/2023

### Hadoop Installation:

**Aim: Implement word count / frequency programs using MapReduce.**

#### Solution:

```
C:\hadoop-3.3.0\sbin>start-dfs.cmd
```

```
C:\hadoop-3.3.0\sbin>start-yarn.cmd
```

Open a command prompt as administrator and run the following command to create an input and output folder on the Hadoop file system, to which we will be moving the sample.txt file for our analysis.

```
C:\hadoop-3.3.0\bin>cd\
```

```
C:\>hadoop dfsadmin -safemode leave
```

DEPRECATED: Use of this script to execute hdfs command is deprecated.

Instead use the hdfs command for it.

Safe mode is OFF

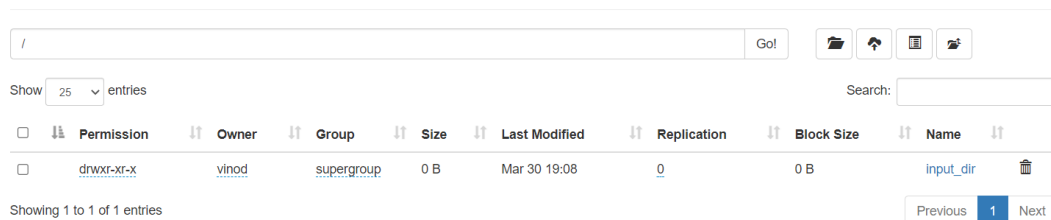
```
C:\>hadoop fs -mkdir /input_dir
```

Check it by giving the following URL at browser

<http://localhost:9870>

Utilities -> browse the file system

#### Browse Directory



The screenshot shows the HDFS web interface at localhost:9870. The address bar shows the path '/'. Below the address bar, there are icons for file operations. A table lists the contents of the directory. The table has columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, Name, and a trash icon. One entry is visible: 'input\_dir' with permissions 'drwxr-xr-x', owner 'vinod', group 'supergroup', size '0 B', last modified 'Mar 30 19:08', and replication '0'. At the bottom, it says 'Showing 1 to 1 of 1 entries' and has 'Previous', '1', and 'Next' buttons.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	vinod	supergroup	0 B	Mar 30 19:08	0	0 B	input_dir

Copy the input text file named input\_file.txt in the input directory (input\_dir) of HDFS.

Make a file in c:\input\_file.txt and write following content in it.

Hadoop Window version is easy compared to Ubuntu version

Now apply the following command at c:\>

```
C:\>hadoop fs -put C:/input_file.txt /input_dir
```

## Browse Directory

/input\_dir

Go!

Show

25

entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		-rw-r--r--		vinod		supergroup		57 B		Mar 30 19:11		1		128 MB		input_file.txt	

Showing 1 to 1 of 1 entries

Previous

1

Next

Verify input\_file.txt available in HDFS input directory (input\_dir).

C:\>Hadoop fs -ls /input\_dir/

```
C:\>hadoop fs -put C:/input_file.txt /input_dir

C:\>hadoop fs -ls /input_dir/
Found 1 items
-rw-r--r--  1 vinod supergroup      57 2023-03-30 19:11 /input_dir/input_file.txt

C:\>
```

Verify content of the copied file

C:\>hadoop dfs -cat /input\_dir/input\_file.txt

You can see the file content displayed on the CMD.

```
C:\>hadoop dfs -cat /input_dir/input_file.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Hadoop Window version is easy compared to Ubuntu version.

C:\>
```

Run MapReduceClient.jar and also provide input and out directories.

C:\>hadoop jar C:/hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar wordcount /input\_dir /output\_dir

```

Reduce input groups=8
Reduce shuffle bytes=103
Reduce input records=8
Reduce output records=8
Spilled Records=16
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=70
CPU time spent (ms)=219
Physical memory (bytes) snapshot=517128192
Virtual memory (bytes) snapshot=792633344
Total committed heap usage (bytes)=392691712
Peak Map Physical memory (bytes)=314761216
Peak Map Virtual memory (bytes)=465485824
Peak Reduce Physical memory (bytes)=202366976
Peak Reduce Virtual memory (bytes)=327180288
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=56
File Output Format Counters
  Bytes Written=65

```

In case, there is some error in executing then copy the file MapReduceClient.jar in C:\ and run the program with the jar file using existing MapReduceClient.jar file as:

*C:\> hadoop jar C:/MapReduceClient.jar wordcount /input\_dir /output\_dir*

Now, check the output\_dir on browser as follows:

## Browse Directory

/

Go!

Show

25

entries

Search:

<input type="checkbox"/>	<div><div></div></div> Permission	<div><div></div></div> Owner	<div><div></div></div> Group	<div><div></div></div> Size	<div><div></div></div> Last Modified	<div><div></div></div> Replication	<div><div></div></div> Block Size	<div><div></div></div> Name	<div><div></div></div>
<input type="checkbox"/>	drwxr-xr-x	<a href="#">vinod</a>	<a href="#">supergroup</a>	0 B	Mar 30 19:29	<a href="#">0</a>	0 B	<a href="#">input_dir</a>	<div><div></div></div>
<input type="checkbox"/>	drwxr-xr-x	vinod	supergroup	0 B	Mar 30 19:30	0	0 B	<a href="#">output_dir</a>	<div><div></div></div>

Click on output\_dir → part-r-00000 → Head the file (first 32 K) and check the file content as the output.

## File contents

```

Hadoop 1
Ubuntu 1
Window 1
compared 1
easy 1
is 1
to 1
version 2

```

Alternatively, you may type the following command on CMD window as:

```
C:\> hadoop dfs -cat /output_dir/*
```

You can get the following output

```
C:\Windows\System32>hadoop dfs -cat /output_dir/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Hadoop 1
Ubuntu 1
Window 1
compared 1
easy 1
is 1
to 1
version 2
C:\Windows\System32>
```

## Output:

File Edit Format View Help

ninad  
shubham  
vighnesh  
ajay  
shubham  
vighnesh

```

Microsoft Windows [Version 10.0.22621.1413]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd C:\Hadoop-3.3.0\sbin

C:\hadoop-3.3.0\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\hadoop-3.3.0\sbin>jps
1584 DataNode
11028 Jps
6952 NodeManager
6476 NameNode
9308 ResourceManager

C:\hadoop-3.3.0\sbin>cd\

C:\>hadoop dfsadmin -safemode leave
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Safe mode is OFF

C:\>hadoop fs -mkdir /input_dir

C:\>hadoop fs -put C:/input_file.txt /input_dir
put: `/input_file.txt': No such file or directory

C:\>hadoop fs -put C:/ninad.txt /input_dir

C:\>hadoop jar C:/hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar wo

```

/

Go!

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	<a href="#">drwxr-xr-x</a>	<a href="#">admin</a>	<a href="#">supergroup</a>	0 B	Apr 25 07:45	<a href="#">0</a>	0 B	<a href="#">input_dir</a>
<input type="checkbox"/>	<a href="#">drwxr-xr-x</a>	<a href="#">admin</a>	<a href="#">supergroup</a>	0 B	Apr 25 07:45	<a href="#">0</a>	0 B	<a href="#">output_dir</a>
<input type="checkbox"/>	<a href="#">drwx-----</a>	<a href="#">admin</a>	<a href="#">supergroup</a>	0 B	Apr 25 07:50	<a href="#">0</a>	0 B	<a href="#">tmp</a>

Showing 1 to 3 of 3 entries

Previous

1

Next

Hadoop, 2020.

## Output:

# Browse Directory

Show  entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">admin</a>	<a href="#">supergroup</a>	0 B	Apr 25 07:50	<a href="#">1</a>	128 MB	<a href="#">_SUCCESS</a>	
<input type="checkbox"/>	-rw-r--r--	<a href="#">admin</a>	<a href="#">supergroup</a>	49 B	Apr 25 07:50	<a href="#">1</a>	128 MB	<a href="#">part-r-00000</a>	

Showing 1 to 2 of 2 entries

Hadoop, 2020.

```
C:\Windows\System32>hadoop dfs -cat /output_dir/*
DEPRECATED: Use of this script to execute hdfs command is deprecated
Instead use the hdfs command for it

ninad 1
shubham 2
vighnesh 2
ajay 1
```

File Edit Format View Help

```
ninad 1
shubham 2
vighnesh 2
ajay 1
```