Hadoop - Map Reduce

Demonstrate the MAP REDUCE programming model by counting the number of words in a file.

Aim:

To demonstrate the Map Reduce programming model by counting the number of words in a file by implementing the procedure to interact with Hadoop API for Accessing HDFS from local file system.

Procedure:

- 1. Open Command prompt and run as administrator and start the Hadoop service.
- 2. Check if the namenode is empty, type "hdfs namenode -format" if it is empty else skip this step.
- 3. Create a new directory in the Hadoop file system using the command "hdfs dfs -mkdir /directory name".
- 4. Upload the input text file to the new directory using the command "hdfs dfs -put C:\input.txt /word count"
- 5. Create the mapper and reducer python programs separately for this program.
- 6. For Hadoop Streaming execute the following command hadoop jar C:\hadoop-3.3.6\share\hadoop\tools\lib\hadoop-streaming-3.3.6.jar ^ -mapper "python C:/mapper.py" ^ -reducer "python C:/reducer.py" ^ -input /word count/input.txt ^ -output /word count/hadoop output/output

Program:

```
mapper.py
#! /usr/bin/env python
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/env python
import sys
prev_word = None
prev_count = 0
for line in sys.stdin:
  line = line.strip()
  word, count = line.split('\t')
  count = int(count)
  if prev_word == word:
    prev_count += count
  else:
    if prev_word:
      print('%s\t%s' %(prev_word, prev_count))
    prev_count = count
    prev_word = word
if prev_word == word:
```

print('%s\t%s' %(prev_word, prev_count))

Output Screenshots:

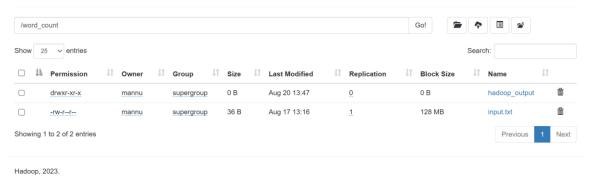
```
C:\Windows\System32>start-dfs
C:\Windows\System32>start-yarn
starting yarn daemons
C:\Windows\System32>jps
22704 NameNode
10116 DataNode
20788 NodeManager
27108 ResourceManager
4248 Jps
```

C:\hadoop-3.3.6\sbin>hadoop fs -mkdir /word_count

C:\hadoop-3.3.6\sbin>hdfs dfs -put C:\input.txt /word_count_

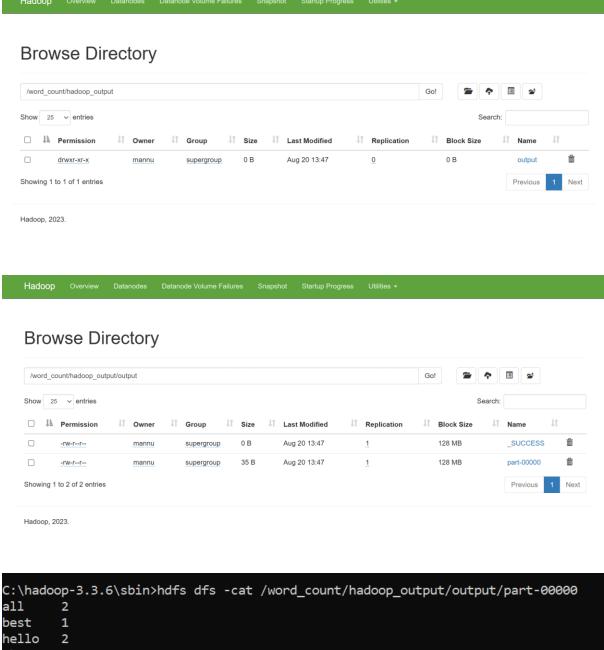
Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▼

Browse Directory



hello all hello world all the best

input.txt



all 2 best 1 hello 2 the 1 world 1 C:\hadoop-3.3.6\sbin>_

Result:

Thus the demonstration of Map Reduce programming model by counting the number of words in a file has been executed successfully.