

Roll No: 20BCE057

Name: Devasy Patel

Course Code: 2CS702

Practical No: 6

Aim: Configure the number of mappers and reducers to execute practical 4 and 5. Evaluate the results and prepare a report on performance

Output:

Mapreudcer function for prac-4

```
PS D:\bdaprac_4> hdfs dfs -put D:\bdaprac_4\input.txt /input4
2023-10-27 09:23:27,545 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
PS D:\bdaprac_4> hdfs dfs -ls /input4/
Found 1 items
-rw-r--r-- 1 Ayush supergroup 44 2023-10-27 09:23 /input4/input.txt
PS D:\bdaprac_4> hadoop jar phase.jar PhraseFrequencyMapReduce /input4/input.txt /output4
2023-10-27 09:27:49,925 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-27 09:27:50,611 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-27 09:27:50,625 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Ayush/.staging/job_1698378600950_0001
2023-10-27 09:27:50,733 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 09:27:50,840 INFO input.FileInputFormat: Total input files to process : 1
2023-10-27 09:27:50,905 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 09:27:50,937 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 09:27:50,957 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-27 09:27:51,060 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 09:27:51,080 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698378600950_0001
2023-10-27 09:27:51,081 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-27 09:27:51,254 INFO conf.configuration: resource-types.xml not found
2023-10-27 09:27:51,255 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-27 09:27:51,672 INFO impl.YarnClientImpl: Submitted application application_1698378600950_0001
2023-10-27 09:27:51,727 INFO mapreduce.Job: The url to track the job: http://LAPTOP-HBK2URJUV:8088/proxy/application_1698378600950_0001/
2023-10-27 09:27:51,728 INFO mapreduce.Job: Running job: job_1698378600950_0001
2023-10-27 09:27:56,808 INFO mapreduce.Job: Job job_1698378600950_0001 running in uber mode : false
2023-10-27 09:27:56,809 INFO mapreduce.Job: map 0% reduce 0%
2023-10-27 09:27:56,828 INFO mapreduce.Job: Job job_1698378600950_0001 failed with state FAILED due to: Application application_1698378600950_0001 failed 2 times due to AM Cont
```

```
starting yarn daemons
D:\hadoop-3.2.1>hadoop jar phase.jar PhraseFrequencyMapReduce /input4/input.txt /output4
JAR does not exist or is not a normal file: D:\hadoop-3.2.1\phase.jar
D:\hadoop-3.2.1>hdfs dfs -ls /
Found 4 items
drwxr-xr-x - Ayush supergroup 0 2023-10-27 09:23 /input4
drwxr-xr-x - Ayush supergroup 0 2023-10-27 09:33 /output4
drwx----- - Ayush supergroup 0 2023-10-27 09:27 /tmp
drwxr-xr-x - Ayush supergroup 0 2023-08-25 09:41 /user
D:\hadoop-3.2.1>hdfs dfs -ls /output4/
Found 2 items
-rw-r--r-- 1 Ayush supergroup 0 2023-10-27 09:33 /output4/_SUCCESS
-rw-r--r-- 1 Ayush supergroup 46 2023-10-27 09:33 /output4/part-r-00000
D:\hadoop-3.2.1>hdfs dfs -cat /output4/part-r-00000
2023-10-27 09:34:31,520 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
ACP 1
DCP 1
DIG 1
DSP 1
IAS 1
IG 2
IPS 1
PC 1
D:\hadoop-3.2.1>
```

Mapreducer function for prac-5

```

PS D:\bdaprac_5> C:\Program\Java\jdk-1.8\bin\javac -classpath $(hadoop classpath) -d . SumDriver.java SumMapper.java SumReducer.java
PS D:\bdaprac_5> jar cvf sum.jar *.class
added manifest
adding: SumDriver.class(in = 1356) (out= 750)(deflated 44%)
adding: SumMapper.class(in = 1738) (out= 709)(deflated 59%)
adding: SumReducer.class(in = 1695) (out= 715)(deflated 57%)
PS D:\bdaprac_5> hdfs dfs -mkdir /input5
PS D:\bdaprac_5> hdfs dfs -put D:\bdaprac_5\input.txt /input5
2023-10-27 10:00:41,504 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
PS D:\bdaprac_5> hadoop jar sum.jar SumDriver /input5/input.txt /output5
2023-10-27 10:01:16,528 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-27 10:01:17,240 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-27 10:01:17,263 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Ayush/.staging/job_1698379320721_0002
2023-10-27 10:01:17,352 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,461 INFO input.FileInputFormat: Total input files to process : 1
2023-10-27 10:01:17,509 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,549 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,567 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-27 10:01:17,709 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,733 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698379320721_0002
2023-10-27 10:01:17,733 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-27 10:01:17,932 INFO conf.Configuration: resource-types.xml not found
2023-10-27 10:01:17,933 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-27 10:01:18,264 INFO impl.YarnClientImpl: Submitted application application_1698379320721_0002
2023-10-27 10:01:18,311 INFO mapreduce.Job: The url to track the job: http://LAPTOP-HBK2URUV:8088/proxy/application_1698379320721_0002/
2023-10-27 10:01:18,312 INFO mapreduce.Job: Running job: job_1698379320721_0002
2023-10-27 10:01:25,459 INFO mapreduce.Job: Job job_1698379320721_0002 running in uber mode : false
2023-10-27 10:01:25,460 INFO mapreduce.Job: map 0% reduce 0%
2023-10-27 10:01:30,574 INFO mapreduce.Job: map 100% reduce 0%

```

```

2023-10-27 10:01:35,635 INFO mapreduce.Job: map 100% reduce 100%
2023-10-27 10:01:35,641 INFO mapreduce.Job: Job job_1698379320721_0002 completed successfully
2023-10-27 10:01:35,706 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=90
    FILE: Number of bytes written=453689
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=123
    HDFS: Number of bytes written=9
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=2838
    Total time spent by all reduces in occupied slots (ms)=2949
    Total time spent by all map tasks (ms)=2838
    Total time spent by all reduce tasks (ms)=2949
    Total vcore-milliseconds taken by all map tasks=2838
    Total vcore-milliseconds taken by all reduce tasks=2949
    Total megabyte-milliseconds taken by all map tasks=2906112

```

```

    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=90
    Reduce input records=6
    Reduce output records=1
    Spilled Records=12
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=94
    CPU time spent (ms)=0
    Physical memory (bytes) snapshot=0
    Virtual memory (bytes) snapshot=0
    Total committed heap usage (bytes)=295698432
  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=21
  File Output Format Counters
    Bytes Written=9
PS D:\bdaprac_5> hdfs dfs -cat /output5/part-r-00000
sum 4641
PS D:\bdaprac_5> 

```