

Name : Kartavya Mandani

Roll No.: 20BCE120

Course: 2CS702 Big Data Analytics

Practical No.: 6 (Configure the number of mappers and reducers to execute practical 4 and 5. Evaluate the results and prepare a report on performance.)

- Configurations in Practical 4:

```
PS C:\Program Files\Java\jdk1.8.0_291\bin&gt; hadoop jar phrase-frequency.jar PhraseFrequencyMapReduce /input/input.txt /output/ mapred.map.tasks=10 -D mapred.reduce.tasks=5
2023-10-13 09:57:21,965 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-13 09:57:22,423 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-13 09:57:22,423 INFO mapreduce.JobResourceUploader: Erasure Coding configuration for path: /tmp/hadoop-yarn/staging/dhruv.staging/job_1697168436407_0009
2023-10-13 09:57:22,505 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:57:22,587 INFO input.FileInputFormat: Total input files to process : 1
2023-10-13 09:57:22,622 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:57:22,645 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:57:22,655 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-13 09:57:22,665 INFO mapreduce.Job: JAR file does not exist: file:/C:/Users/dhruv/Desktop/phrase-frequency.jar
2023-10-13 09:57:22,675 INFO mapreduce.Job: User specified no jobconf properties.
2023-10-13 09:57:22,675 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168436407_0009
2023-10-13 09:57:22,766 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-13 09:57:22,902 INFO conf.Configuration: resource-types.xml not found
2023-10-13 09:57:22,904 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-13 09:57:22,928 INFO impl.YarnClientImpl: Submitted application application_1697168436407_0009
2023-10-13 09:57:22,988 INFO mapreduce.Job: The url to track the job: http://DESKTOP-365G0V4:8088/proxy/application_1697168436407_0009
2023-10-13 09:57:22,998 INFO mapreduce.Job: Running job: job_1697168436407_0009
2023-10-13 09:57:30,149 INFO mapreduce.Job: map 0% reduce 0%
2023-10-13 09:57:34,438 INFO mapreduce.Job: map 100% reduce 0%
2023-10-13 09:57:39,009 INFO mapreduce.Job: map 100% reduce 100%
2023-10-13 09:57:39,227 INFO mapreduce.Job: Job job_1697168436407_0009 completed successfully
2023-10-13 09:57:39,405 INFO mapreduce.Job: Counters: 32
  File System Counters
    FILE: Number of bytes read=55
    FILE: Number of bytes written=453653
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=133
    HDFS: Number of bytes written=33
    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

```

```
PS C:\Program Files\Java\jdk1.8.0_291\bin&gt; hadoop jar phrase-frequency.jar PhraseFrequencyMapReduce /input/input.txt /output/ mapred.map.tasks=5 -D mapred.reduce.tasks=2
2023-10-13 09:56:35,260 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-13 09:56:35,763 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-13 09:56:35,763 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dhruv.staging/job_1697168436407_0008
2023-10-13 09:56:35,867 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:56:35,994 INFO input.FileInputFormat: Total input files to process : 1
2023-10-13 09:56:36,016 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:56:36,441 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-13 09:56:36,545 INFO ssl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:56:36,558 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168436407_0008
2023-10-13 09:56:36,706 INFO conf.Configuration: resource-types.xml not found
2023-10-13 09:56:36,707 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-13 09:56:36,766 INFO impl.YarnClientImpl: Submitted application application_1697168436407_0008
2023-10-13 09:56:36,802 INFO mapreduce.Job: The url to track the job: http://DESKTOP-365G0V4:8088/proxy/application_1697168436407_0008
2023-10-13 09:56:43,964 INFO mapreduce.Job: Running job: job_1697168436407_0008
2023-10-13 09:56:43,968 INFO mapreduce.Job: map 0% reduce 0%
2023-10-13 09:56:48,066 INFO mapreduce.Job: map 100% reduce 0%
2023-10-13 09:56:51,335 INFO mapreduce.Job: map 100% reduce 100%
2023-10-13 09:56:51,145 INFO mapreduce.Job: Job job_1697168436407_0008 completed successfully
2023-10-13 09:56:53,223 INFO mapreduce.Job: Counters: 52
  File System Counters
    FILE: Number of bytes read=55
    FILE: Number of bytes written=453653
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=133
    HDFS: Number of bytes written=33
    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)=2459
    Total time spent by all reduces in occupied slots (ms)=2537
    Total time spent by all map tasks (ms)=2459
    Total time spent by all reduce tasks (ms)=2537
    Total vcore-milliseconds taken by all map tasks=2459
    Total vcore-milliseconds taken by all reduce tasks=2537
    Total megabyte-milliseconds taken by all map tasks=2518016
    Total megabyte-milliseconds taken by all reduce tasks=2597888

```

- Comparison:

configuration	case1	case2
Mapred.map.task	10	5
Mapred.red.task	5	2
Time Taken by mapper(milisec)	2485614	2518016
Time Taken by reducer(milisec)	2284983	2597888

- Configurations in Practical 5:

```
PS C:\Program Files\Java\jdk1.8.0_291\bin\$> hadoop jar sumDriver.jar SumDriver mapred.map.tasks=5 -D mapred.reduce.tasks=2
2023-10-13 09:50:04,584 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-13 09:50:05,078 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-13 09:50:05,089 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dhruv/.staging/job_1697168436407_0005
2023-10-13 09:50:05,169 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:50:05,276 INFO input.FileInputFormat: Total input files to process : 1
2023-10-13 09:50:05,362 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:50:05,392 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:50:05,404 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-13 09:50:05,430 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:50:05,440 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168436407_0005
2023-10-13 09:50:05,448 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168436407_0005
2023-10-13 09:50:05,448 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-13 09:50:05,460 INFO conf.Configuration: resource-types.xml not found.
2023-10-13 09:50:05,602 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-13 09:50:05,647 INFO impl.YarnClientImpl: Submitted application:application_1697168436407_0005
2023-10-13 09:50:05,678 INFO mapreduce.Job: The user can track the job: http://DESKTOP-3656BV4:8088/proxy/application_1697168436407_0005/
2023-10-13 09:50:05,679 INFO mapreduce.Job: Running job: job_1697168436407_0005
2023-10-13 09:50:12,853 INFO mapreduce.Job: Job job_1697168436407_0005 running in uber mode : false
2023-10-13 09:50:12,854 INFO mapreduce.Job: map 0% reduce 0%
2023-10-13 09:50:18,942 INFO mapreduce.Job: map 100% reduce 0%
2023-10-13 09:50:24,002 INFO mapreduce.Job: map 100% reduce 100%
2023-10-13 09:50:25,027 INFO mapreduce.Job: Job job_1697168436407_0005 completed successfully
2023-10-13 09:50:25,111 INFO mapreduce.Job: Counters: 50
    File System Counters
        FILE: Number of bytes read=90
        FILE: Number of bytes written=453697
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=119
        HDFS: Number of bytes written=7
        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)=2855
        Total time spent by all reduces in occupied slots (ms)=2758
        Total time spent by all map tasks (ms)=2855
        Total time spent by all reduce tasks (ms)=2758
        Total vcore-milliseconds taken by all map tasks=2855
        Total vcore-milliseconds taken by all reduce tasks=2758
        Total megabyte-milliseconds taken by all map tasks=2923520
        Total megabyte-milliseconds taken by all reduce tasks=2824192
```

```

PS C:\Program Files\Java\jdk1.8.0_291\bin> hadoop jar sumDriver.jar mapred.map.tasks=10 -D mapred.reduce.tasks=5
2023-10-13 09:49:12,874 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-13 09:49:13,309 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-13 09:49:13,387 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dhruv-staging/job_1697168436407_0004
2023-10-13 09:49:13,401 INFO input.FileInputFormat: Total input files to process : 1
2023-10-13 09:49:13,573 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:49:13,592 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:49:13,604 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-13 09:49:13,702 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-13 09:49:13,707 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168436407_0004
2023-10-13 09:49:13,708 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-13 09:49:13,846 INFO conf.Configuration: resource-types.xml not found
2023-10-13 09:49:13,904 INFO impl.YarnClientImpl: Submitted application application_1697168436407_0004
2023-10-13 09:49:13,944 INFO mapreduce.Job: The url to track the job: http://DESKTOP-365G6V4:8088/proxy/application_1697168436407_0004/
2023-10-13 09:49:13,944 INFO mapreduce.Job: Job: job_1697168436407_0004
2023-10-13 09:49:20,103 INFO mapreduce.Job: Job job_1697168436407_0004 running in uber mode : false
2023-10-13 09:49:20,108 INFO mapreduce.Job: map 0% reduce 0%
2023-10-13 09:49:25,231 INFO mapreduce.Job: map 100% reduce 0%
2023-10-13 09:49:30,298 INFO mapreduce.Job: map 100% reduce 100%
2023-10-13 09:49:30,305 INFO mapreduce.Job: Job job_1697168436407_0004 completed successfully
2023-10-13 09:49:30,305 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=90
  FILE: Number of bytes written=453697
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=119
  HDFS: Number of bytes written=7
  HDFS: Number of read operations=8
  HDFS: Number of large read operations=2
  HDFS: Number of write operations=2
  HDFS: Number of byte 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)=2688
  Total time spent by all reduces in occupied slots (ms)=2648
  Total time spent by all map tasks (ms)=2688
  Total time spent by all reduce tasks (ms)=2648
  Total vcore-milliseconds taken by all map tasks=2688
  Total vcore-milliseconds taken by all reduce tasks=2648
  Total megabyte-milliseconds taken by all map tasks=2752512
  Total megabyte-milliseconds taken by all reduce tasks=2711552

```

- Comparison:

configuration	case1	case2
Mapred.map.task	10	5
Mapred.red.task	5	2
Time Taken by mapper(milisec)	2752512	2923520
Time Taken by reducer(milisec)	2711552	2824192

THE END

:)