

Practical-6

Name : Yash Marthak
Roll No. : 19BCE122
Couse Name : Big Data Analytics

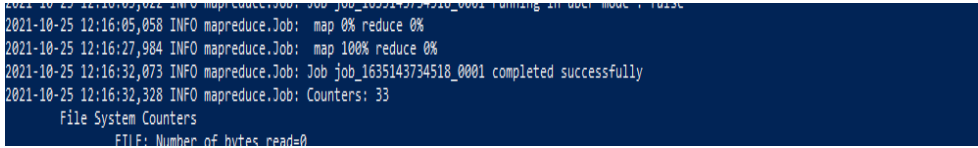
AIM: Analysis impact of different number of mapper and reducer on same definition of practical 4.

Method 1:

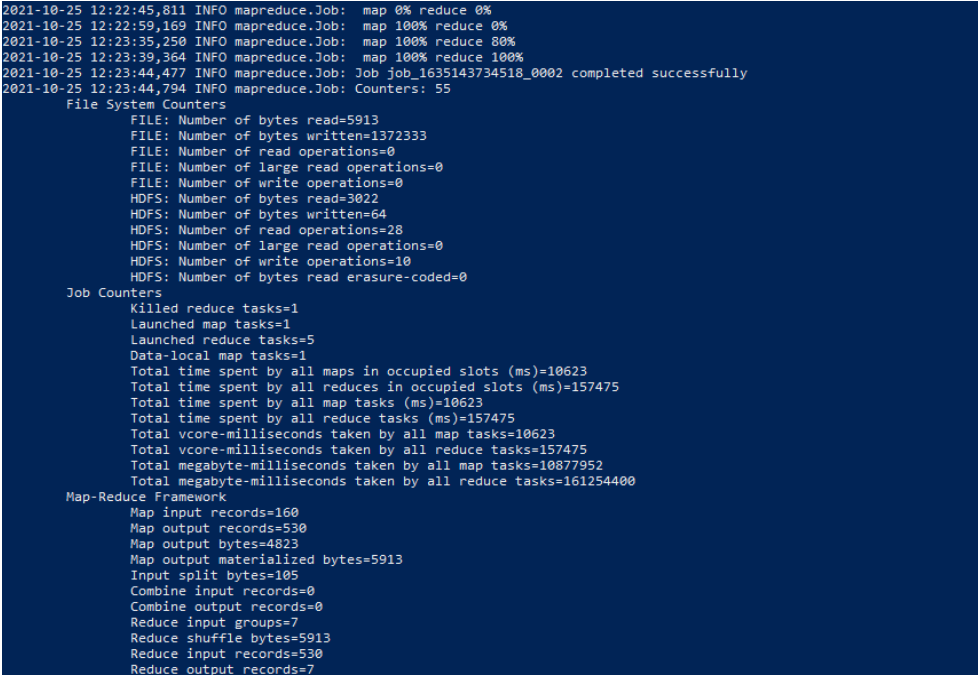
Changes in the code:

1. `job.setNumReduceTasks(0);`

- a. Here, number of reducer are taken as 0.
There is no need of reducer as
They are used for data consolidsiotn,
aggregation rather than heavy computation.
- b. In this case output will be generated by
mappers only which is know
as map-only job in Hadoop.

c. The screenshot shows Hadoop job logs for a map-only job. It includes timestamps, log levels (INFO), and progress reports for map and reduce tasks. The job is identified as 'Job job_1635143734518_0001' and is marked as 'completed successfully'. The counters section shows 33 map tasks and 0 reduce tasks.

2. `job.setNumReduceTasks(5);`

The screenshot shows Hadoop job logs for a job with 5 reducers. It includes timestamps, log levels (INFO), and progress reports for map and reduce tasks. The job is identified as 'Job job_1635143734518_0002' and is marked as 'completed successfully'. The counters section shows 5 map tasks and 5 reduce tasks. The 'Map-Reduce Framework' section provides detailed statistics on input/output records, bytes, and time spent by mappers and reducers.

Method 2:

Using Command Line Arguments

```
hadoop jar -D mapreduce.job.reduces=0  
Lab_6_MapReduceProgram.jar WordCount \input  
\out
```

put2

Conc

lu

sion

Generally, number of reducers should be one of these:

- A multiple of the block size
- The one which take less time to run task
- Create fewest files possible

Pros of more reducer: The real work is done faster

Cons of more reducer: Slower start-up time, more network and disk overhead, throughout and aftercomputation.

Hence, we can conclude that don't choose too fewreducers as if they are very big and one fails it'lltake much more time to recover. Don't choose too many reducers because if the partition issingle thenall the work will be done by single reducer despite having a greater number of reducers. Hence, this decision is combined with number of partitions.