Practical Evaluation

Shadi Ibrahim

Deadline for sending the report is March 24th 11:59PM

The report should be constructed by two students. Reports (in PDF) should be sent by email to *shadi.ibrahim@inria.fr*. A constructed answer of each question is expected: discuss why you conduct this experiment and explain the results and your observations. It will be great if you can discuss also the experimental setup and the metrics for each question.

1.: Mitigating stragglers/failures in Hadoop

Question 1.1 Use data set of 20GB and run sort benchmark using block size (i.e., 128MB) and the default replication factor (3) with speculation enabled and disabled (try to log-in one node and write a program to overload the CPU and/or the disk). Set the map/reduce slots number to 8.
See the execution times, the number of speculative copies, and average (maximum) task runtimes for both map and reduce tasks, what can you observe?

Question 1.2 Use data set of 20GB and run sort benchmark using block size (i.e., 128MB) and the default replication factor (3). For each run, set different expiry interval (30seconds, and 1 minutes). Try to stop the tasktracker daemon on one node (before the completion of the map tasks and after the completion of map tasks).
See the execution times, the number of killed tasks, what can you observe?

2.: Application specific optimization in Hadoop

Question 2.1 Use the Data set 10GB and run wordcount and sort benchmarks using different values (i.e., 0.05, 0.5, and 1). See the execution time, what can you observe?

3.: Scheduling in Hadoop "Bonus"

Question 3.1 Run 6 sort applications (vary the size from 3*2GB, 2*4GB 1*10GB) under (1) Fifo scheduler, (2) Fair scheduler (preemption disabled), and (3) Fair scheduler (preemption enabled). For Fair scheduler, use fair allocation within the pool – you are using one pool.
Jobs are launched with a decreasing input size order and 10 second interval between two jobs.
See the execution time, data locality and the waiting time of the applications. what can you observe?