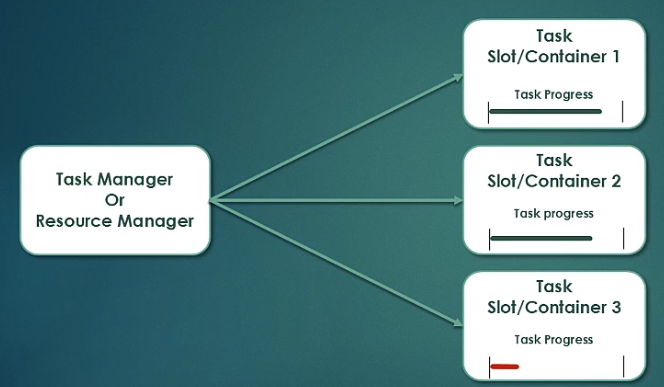
**PERFORMANCE TUNNING FEATURES**

1. Speculative Executive :::: Advantage of Speculative execution in Hadoop is --- That it divides a Task into multiple small Task and Process them in Parallel.

Compare to do the same task on a single node , Hadoop Parallel processing has Advantage of HIGH THROUGHPUT.

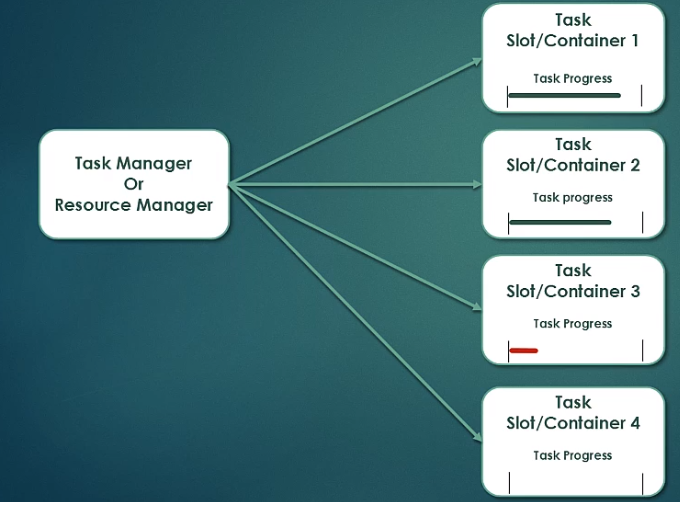
\*\*\*\* But in Case, If any of its Task performs poorly the performance of whole job will goes Down….

EX:::::



Here we can see that Task 3 is following behind the rest … This can be done due to Hardware degradation or Software Misconfiguration As well..

\*\*\* In this Case Hadoop RE launches the Task to another Machine…



\*\*\* Now Task which was running in Container 3 is Now Running on 2 Different Machine ( Container 3 & Container 4 )… Now Out of theses 2 Container who will finish the Task faster that output will be taken and Other Task will be KILLED.

**Key Points :::**

1. Speculative Task are Launched only after all the Task of a JOB has been Launched…
2. JOB TRACKER then Monitors IF there is any Task Falling Behind. And then Only It would speculatively RE execute the Task .
3. Second It is an Optimization Feature not a Reliability Feature. What it implies that If the Task is running Slow because of BUGS in Code , Heap won’t be able to fix it or Diagnose it or PIN POINT error on the Code..

IT is just ensuring that Underlined Hardware & Software Configurations are not the Reason for Slow Progress of the TASK. And So It would TRY to run it on the Other Node .

So , The whole Job Finishes as quickly as Possible. Either of Original Task or Speculative Task can Finish First.

\*\* As soon as One Task Finished other one Will be Killed.

\*\*\* Property for Speculative Execution are set by

**mapred.map.task.speculative.execution**

**mapred.reduce.task.speculative.execution**

\*\* These are BOLEAN property by Default set to TRUE. This means that Speculative execution is enable by default.

\*\* These Property are recommended to be set to be true.

\*\*\*\*\*\*\*\*\*\*\*But WE can Set it as FALSE also. This can be in the case If our Cluster is already overloaded and We need not Overload the resources with the speculative Task…

\*\* Some Installation Prefer NO to SPECULATIVE INSTALLATION ON REDUCE SIDE--- Because to start another copy of Reduce Task- Output of Map Task has to be fetched from the Network -- Which Increase LOAD on NETWORK.