

Is a Slow Mapper a Bottleneck?

- **It is possible for one Map task to run more slowly than the others**
 - Perhaps due to faulty hardware, or just a very slow machine
- **It would appear that this would create a bottleneck**
 - The `reduce` method in the Reducer cannot start until every Mapper has finished
- **Hadoop uses *speculative execution* to mitigate against this**
 - If a Mapper appears to be running significantly more slowly than the others, a new instance of the Mapper will be started on another machine, operating on the same data
 - A new *task attempt* for the same task
 - The results of the first Mapper to finish will be used
 - Hadoop will kill off the Mapper which is still running