

Is Shuffle and Sort a Bottleneck?

- **It appears that the shuffle and sort phase is a bottleneck**
 - The `reduce` method in the Reducers cannot start until all Mappers have finished
- **In practice, Hadoop will start to transfer data from Mappers to Reducers as soon as the Mappers finish work**
 - This avoids a huge amount of data transfer starting as soon as the last Mapper finishes
 - The `reduce` method still does not start until all intermediate data has been transferred and sorted