

Map Reduce

1. What is the functionality of Mapper, Combiner, Partitioner and Reducer?
2. What is the main difference of Map Reduce and Yarn?
3. What are all reasons behind the failure of Map Reduce?
4. What is the flow of Yarn Scheduler?
5. What is the difference between Combiner and Reducer?
6. How many default Mapper and Reducer will run for Single Map Reduce job?
7. What is Speculative execution in hadoop?
8. What is "Distributed Cache" in a Map Reduce Framework and its why we need it?
9. What is the default value of map and reduce max attempts?
10. What is the InputFormat and OutputFormat of mapper and reducer?
11. What is shuffling and sorting in Map Reduce?
12. What happen if number of reducers is set to 0 in Hadoop?
13. Why aggregation cannot be done in Mapper?
14. What is Record Reader and Record Writer in map reduce?
15. What are intermediate results in Hadoop? Where hadoop write those result sets?
16. Are Job Tracker and Task Tracker present on the same machine?
17. Difference between MR and Tez?
18. If a Data Node is marked as decommissioned, can it be chosen for replica placement?
19. How will you decommission the data node? What are all the checks needs to do?