Name : MAULIK VIPINCHANDRA DHAMECHA
Enrollment No : 159997107017
Branch : Computer/ IT Engineering
Title of the Thesis : IMPROVED MAPREDUCE ALGORITHM WITH EFFICIENCY IN STRUCTURE AND RELATIONS IN THE LARGE BODY OF THE DATA (BIGDATA)

Abstract

"Big Data" refers to enormous amounts of unstructured data produced by high performance applications falling in a wide and heterogeneous family of application scenarios: from scientific computing applications to social networks, from government applications to medical information systems, and so forth. Recent research in big data has shown that the amount of data continues to increase at an exponential rate. To process a massive amount of data on computing clusters, you need a powerful computing model like Hadoop and MapReduce. MapReduce framework is a programming model that processes terabytes of data in a very less amount of time. To achieve excellent performance, big data requires proper scheduling. Scheduling Technique is used to reduce starvation, increase the usage of resources, and also to assign jobs for available resources.

For the Hadoop MapReduce model, various scheduling algorithm has been developed, which differs widely in design, behaviour, and handling various issues. First In First Out (FIFO) scheduling is the simplest and most efficient among the scheduling algorithm. A significant drawback of FIFO scheduling is the poor response time for short jobs in comparison to large jobs and low performance in handling multiple types of jobs. Existing resource allocation scheduling does not take the weight of each job into consideration which leads to unbalanced performance among nodes. The objective of this research is to provide a strategy that balances resource allocation between short jobs and long jobs, considering the weight of each job. This study provides a faster response time to smaller jobs to achieve better performance in scheduling.

In the current scenario, there is no consideration of system load during allocation to task trackers and if one task tracker is slow, it can delay the whole MapReduce job. So, for this kind of system, we have to calculate a load of all data nodes as well as the speed of every task tracker. Assign the task to that node whose speed of task tracker is better to compare to other data nodes and maintain the load balancing.

This PhD Thesis would be useful for smooth scheduling in big data analytics.

List of Publications :

1) Maulik Dhamecha, Dr. Tejas Patalia, "Fundamental Survey of Map Reduce in Bigdata with Hadoop Environment", Spinger – CCIS (2018).
2) Maulik Dhamecha, Dr. Tejas Patalia, "Comparative study of Dynamic Load Balancing algorithm in large scale data (Big data)", International Journal of Advanced Science and Technology (2020)
3) Maulik Dhamecha, Dr. Tejas Patalia, "Scheduling issue for Dynamic Load Balancing of mapreduce in large scale data (Big data)", Journal of Xidian University (2020)