Abstract

MapReduce is an emerging programming paradigm for data parallel applications proposed by Google to simplify large-scale data processing. MapReduce implementation consists of map function that processes input key/value pairs to generate intermediate key/value pairs and reduce function that merges and converts intermediate key/value pairs into final results. The reduce function can only start processing after completion of the map function. Due to dependencies between map and reduce function, if the map function is slow for any reason, this will affect the whole running time. In this technique, the message passing interface (MPI) strategies is used to implement MapReduce which reduces the runtime and optimized data...
exchange. MPI is used for algorithmic parallelization. MapReduce with MPI combines redistribution and reduce and moves them into the network. In this paper, new technology used as MapReduce overlapping using MPI, which is an enhancing structure of the MapReduce programming model for fast data processing. This implementation is based on running the map and the reduce functions concurrently in parallel by exchanging partial intermediate data between them in a pipeline fashion using MPI. At the same time, performing the algorithm parallelism in order to increase the performance with data parallelism of using overlapping mapreduce MPI. MPI support more efficiently all MapReduce applications.

References

- Hishan Mohamed, Stephane Marchand-Maillet, "Enhancing MapReduce using MPI and an optimized data exchange policy," in 2012 41st International Conference on Parallel Processing Workshops.

Index Terms

Computer Science Information Sciences

Keywords

Hadoop Mapreduce Overlapping Mpi-mapreduce Parallel Mapreduce.