

IHadoop: Improve MapReduce Performance

Mohammed Qunper, Osama Badawy, Mohammed Kholief College of Computing and Information Technology, Arab Academy for Science, Technology, and Maritime Transport, Alexandria, Egypt m.qunper@gmail.com, obadawy@aast.edu, kholief@aast.edu

Abstract: *The performance of Hadoop dependent on many of points, the data partitioning is one of this point, the node specification is the other point. In the real world, the data is often highly skewed, which may cause losing time for the jobs. In this paper we study the skew problem in reduce and map phases, where map phase need to collect bocks and reduce phase need to collect key groups. We outline our solution to develop map and reduce phases, by decrease preparation Maper and Reducer, so we select locality base partitioning and developed to solve power loss by using node specification to decrease map time, and key group robust to decrease reduce time.*

Keywords: *Hadoop, MapReduce, Big data, Cloud computing, Parallel computing.*
