

Jordan University of Science and Technology

Index-Based Join in MapReduce Using Hadoop MapFiles

Authors: Amer Al-Badarneh, Mohammed Al-Rudaini, Faisal Ali, Hassan Najadat

Abstract: MapReduce stays an important method that deals with semi-structured or unstructured big data files, however, querying data mostly needs a Join procedure to accumulate the desired result from multiple huge files. Indexing in other hand, remains the best way to ease the access to a specific record(s) in a timely manner. In this paper the authors are investigating the performance gain by implementing MapFile indexing and Join algorithms together.