

要 旨

Scan 計算パターンの MapReduce 上での実現に関する研究

畑島 正和

近年，1 台の PC では扱いきれないほどデータは巨大化しており，並列計算の重要性が増加してきている．しかし，並列計算を行うプログラムを作成するのは難しい．これを解決する手法の一つに MapReduce というものがある．MapReduce は，大規模な計算ノード・クラスタ上で小さな処理を並列に実行することにより高速な処理を行える，Google によって開発されたプログラミングモデルである．MapReduce の提案以前より，要素間に順序のあるデータ構造であるリストに対して並列計算の研究が行われており，その中でも reduce，scan という計算パターンが重要とされている．しかし，reduce を MapReduce 上に実装している事例はあるが，scan をどのように実装するかは明らかではない．

本研究では，計算パターン scan を MapReduce 上に実現し，それによる速度/台数効果の検証を行った．結果より，リストに対する並列計算を MapReduce プログラミングに応用できることが確認できた．

キーワード MapReduce リスト 並列計算 reduce scan

Abstract

Realizing Scan Computation Pattern on MapReduce

Masakazu HATASHIMA

In recent years, has become huge data than can be handled in a single PC, the importance of parallel computing has been increasing. However, to create a program that performs parallel computation is difficult. MapReduce is high-speed processing can be performed in parallel by running a small processing on large-scale compute-node cluster, is a programming model has been developed by Google. Than the previous proposal of MapReduce, study of parallel computation has been made to the list which is a data structure in order between the elements, calculated pattarn reduce and scan is also important in that. However, the case has been implemented on the MapReduce to reduce, but how to implement the scan is not clear.

In this study, realized on the scan computation pattern on MapReduce was carried out to verify the speed and volume effect by it. From the results, MapReduce purogramming can be applied to parallel computing for the list could be confirmed.

key words MapReduce list Parallel computation reduce scan