

## 平成 22 年度 情報工学コース卒業研究報告要旨

阿草 研究室	氏 名	村 松 沙 那 恵
卒業研究題目	MapReduce を用いたアプリケーションの設計法	

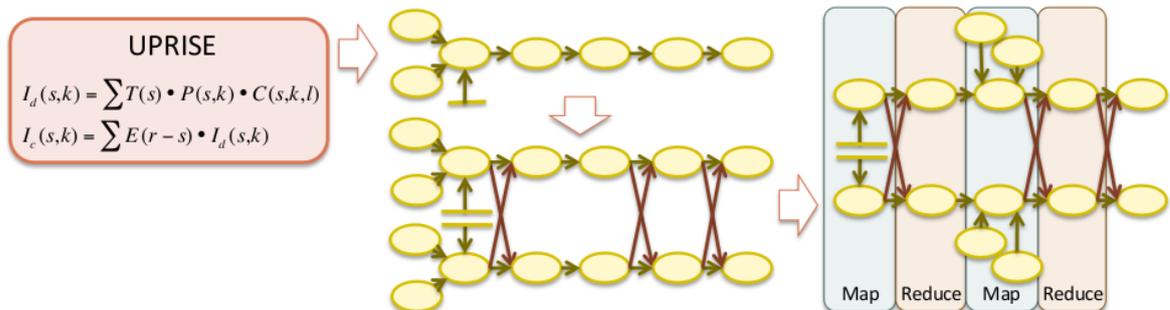
本研究では大規模分散並列のプログラミングモデルである MapReduce をソフトウェア開発に適用する際の設計支援法を提案する。

近年，ブログや SNS など個人の生み出すデータ量が著しく増大している．さらにセンサーやログなど機械から収集されるデータも爆発的に増えている．このような背景から大量のデータを処理する需要が高まっている．これまでシステムの処理性能をあげるには個々の計算機の性能を強化する手法がとられてきた．しかしこの手法では処理しきれないデータ量があり限界を向かえつつある．さらに規模を拡大・縮小するためにマシンの入替を行う手間があるため柔軟に性能をあげることができない．

そこで安価な計算機を多数用いることで規模を拡大していく手法がとれている．これは現在のクラウドコンピューティングを支えている手法であり，多数の計算機を効率よく用いるために様々な技術が適用されている．その中の 1 つである MapReduce は並列分散処理を行うプログラミングモデルである．このモデルを使うことで，並列分散を行ううえの面倒な詳細を気にすることなく map/reduce 処理とその入出力である key/value のペアを決定するだけで自動的に並列分散実行をする．

しかしながら，MapReduce による並列分散処理プログラムは逐次的な処理とのギャップが存在する．どのような処理を map/reduce 処理をするとその利点を生かせるのか，また実装するにはどのようにプログラムを書けばいいかの指針がない．そこで本論文では MapReduce を用いたアプリケーション開発の設計手法を提案する．最初に処理の入出力データと処理内容を左から順に処理が流れるデータフロー図に書く．それを基に map/reduce 処理の可能性の箇所やデータの依存箇所を調べ map/reduce 処理の特定を行う．処理で利用されるデータを調べて key/value の特定を行う．map/reduce の処理内容と key/value を決定することで MapReduce を適用する支援を行う．

本研究では講義動画検索システム UPRISE の Hadoop を用いた実現を対象にした．UPRISE の要求仕様から，データフロー図を記述し，提案手法に基づき map/reduce による実装を行った．



適用実験の結果から，一部制約はあるものの，提案手法に基づき map/reduce 処理を適用する箇所を特定することが確認でき，本研究の有用性を明かにした．