

平成 25 年度図書館情報メディア系プロジェクト研究 研究成果報告書

種 目	外部資金獲得支援・萌芽的研究支援		研究代表者 氏 名	川原崎 雅敏
研究課題	ネットワーク連携型タスクスケジューリングを用いた Hadoop 性能の改善方式			
研究組織（研究代表者及び研究分担者）				
氏 名	所属研究機関・部 局・職	現在の専門	役割分担	
川原崎雅敏	図情メディア系・教授	情報通信ネットワーク	統括	
杉本重雄	図情メディア系・教授	デジタルライブラリ	分散ファイル管理方式	
阪口哲男	図情メディア系・准教授	情報科学	網管理方式	
研究目的				
<p>ビッグデータを分散並列処理するフレームワーク MapReduce とそのオープンソフト実装である Hadoop が世界で広く用いられている。膨大なデータを計算機クラスタで並列処理する Hadoop では、分散並列処理の過程においてマシン間に発生する大量のデータ転送がジョブ処理性能を低下させる事例が報告されているが、そのメカニズムは明確になっていない。本研究では、実機クラスタとクラウド上のクラスタでの実験を通じて、データ転送による性能低下のメカニズムを解明し、それに基づいて計算リソースだけでなくネットワークリソースの利用状況も考慮したタスクスケジューリング方式を提案・評価し、Hadoop 開発サイドにフィードバックすることを目的とする。</p>				
研究成果				
<p>実機クラスタおよびクラウドクラスタでの実験により、ジョブ処理性能の低下はネットワークボトルネックに伴うデータ転送待ちがタスクスロットを無効保留し、有効スロット数を減少させるためであること、また、ハードディスク入出力(HDD IO)の輻輳も同様のメカニズムでジョブ処理性能を低下させることを明らかにした。この分析に基づいて、処理ノードのネットワーク使用率と HDD 使用率を加味してノードにタスクを割り当てるスケジューリング方式を提案し、ジョブ処理時間を短縮できることを示した。さらに、実機クラスタはコスト面で大規模化できず、クラウドクラスタは網トポロジを設定できないという実験環境の問題を克服するために、ネットワークシミュレータ ns-3 を用いた Hadoop 処理の性能解析シミュレータを開発した。シミュレーションで実験環境を構築することでスケール面、ネットワーク設定自由度の面で利点が得られた。</p>				
研究発表・特許等の成果一覧、特記事項等（論文のPDFまたはコピーを添付のこと）				
<ol style="list-style-type: none"> Hyuma Watanabe, Masatoshi Kawarasaki, "Impact of Data Transfer to Hadoop Job Performance: – Architectural Analysis and Experiment –", ACMSE 2014 : 52nd Annual ACM Southeast Conference, 6 pages, March, 2014 渡邊飛雄馬、川原崎雅敏：「データ転送量を考慮した Hadoop スケジューラ改善方式の提案と評価」、信学技報 vol. 113, no. 473, IN2013-173, pp. 175-180, 2014 年 3 月 田中大地、渡邊飛雄馬、川原崎雅敏：「ns-3 を用いた Hadoop 動作解析シミュレータの開発」、信学技報, vol. 113, no. 473, IN2013-174, pp. 181-186, 2014 年 3 月. 				