

Title	A Study on Reliability and Performance Improvement of Distributed System for Large-Scale Data Processing
Author(s)	Kirihata, Yasuhiro
Citation	大阪大学, 2012, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/24736
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

【3】

氏名	桐畑康裕
博士の専攻分野の名称	博士(情報科学)
学位記番号	第25615号
学位授与年月日	平成24年9月25日
学位授与の要件	学位規則第4条第1項該当 情報科学研究科マルチメディア工学専攻
学位論文名	A Study on Reliability and Performance Improvement of Distributed System for Large-Scale Data Processing (大規模データ分散処理システムにおける信頼性及び処理性能向上に関する研究)
論文審査委員	(主査) 教授 薦田 憲久 (副査) 教授 藤原 融 教授 下條 真司 教授 西尾 章治郎 教授 細田 耕

論文内容の要旨

近年、ネットワーク接続されるデバイス端末数の増加や帯域幅の増大に伴い、全世

界で生成・蓄積される電子データが指数関数的に増加している。IDCの調査報告によると、その増加率は2014年には年率60%に達すると言われており、2020年には、全データ量は35.2ゼットバイトに達するとの予測である。こうした著しいデータ量の増大に伴い、低コストで大規模データを格納・高速処理するニーズが高まっている。高性能なコモディティサーバの価格の下落に伴い、こうしたサーバを利用した、大規模データを取り扱うクラスタシステムや分散処理システムの技術に注目が集まっている。このような分散システムにおいて、セキュリティ、処理性能、スケーラビリティ、アベイラビリティが重要な指標である。本論文では、大規模データを取り扱う分散処理システムに関し、以下の課題について論じる。

まず、大規模データを取り扱う分散システムのセキュリティに関して、有効なデータ保護方式の一つとしてシンクライアントがある。しかしながら、導入・運用コストが高い問題から、一般的に導入が進んでおらず、大きな課題である。本課題については、汎用PC端末を活用した低コストな擬似シンクライアントである「WriteShield」を提案する。ボリュームフィルタ層で、ローカルディスクへの保存禁止を実現しており、書込みデータの物理メモリ上へのキャッシュ溢れを防止する仮想キャッシュ機構を実現している。

大規模データの高速処理に関しては、汎用PCクラスタによるバッチ処理技術やリアルタイム処理技術があるが、大規模データのバッチ処理に関して重要な課題の一つに、分散検索システムにおける分散インデックスの高速生成処理がある。インデックス生成において、汎用PCクラスタによる並列分散処理を行うが、その際に必要となる各ファイルと分割インデックス間のマッピングデータの省メモリ化と、検索ノードの追加・削除時のインデックス再構築時間の短縮が重要な課題である。本課題については、従来のコンシステントハッシュ法を改良した、スロット型仮想ノード方式を提案する。

大規模データのリアルタイム処理に関して、代表的なもの一つに、大規模な3次元画像データの高速レンダリング技術があるが、多数ポリゴンで構成される高解像度画像データを取り扱うためのインフラ構成と、クラスタノード間を流れる大量のデータフロー制御が重要な課題である。本課題については、光スイッチベースの汎用PCクラスタによるSort-Lastレンダリング方式を提案する。本方式では、木構造のクラスタ構成を採用することで、光スイッチ内のスイッチング遅延問題を回避する。また、ノード間でのデータ転送に関し、フィードバック制御を行うことで、ノード間のスムーズなデータフローを実現する。

論文審査の結果の要旨

近年の著しいデータ量の増大に伴い、低コストで大規模データを格納・高速処理するニーズが高まっている。高性能な汎用サーバの価格下落に伴い、汎用サーバを利用した、大規模データを取り扱うクラスタシステムや分散処理システムの技術に注目が集まっている。本論文は、大規模データを取り扱う分散処理システムに関し、(1)低コストシンクライアントの実現、(2)分割インデックスの生成・再構築における省リソース化・性能向上、(3)大規模3次元画像データのレンダリング高速化、の3点の課題を踏まえ、大規模データ分散システムにおける信頼性、及び処理性能向上に関する研究成果を纏めたものである。その主要な成果を要約すると次の通りである。

- (1) 大規模データを取り扱う分散システムのセキュリティに関して、汎用PC端末を活用した低コストな擬似シンクライアントである「WriteShield」を提案している。ボリュームフィルタ層で、ローカルディスクへの保存禁止を実現しており、書込みデータの物理メモリ上へのキャッシュ溢れを防止する仮想キャッシュ機構を実現している。実際のI/Oログを使用したシミュレーションの結果、仮想キャッシュ機構のページング処理には、RANDOMが最適であることを確認してい

る。また、ディスクI/O性能は、通常モード時においては物理ディスクのI/O性能より高速であり、ページング処理時においても、実用上問題ない性能であることを示している。

- (2) 大規模データのインデックス生成において、汎用PCクラスタによる並列分散処理を行う際、各ファイルと分割インデックス間のマッピングデータの省メモリ化と、検索ノードの追加・削除時のインデックス再構築時間の短縮が、重要な課題である。本課題について、スロット型仮想ノード方式によるコンシステントハッシュ法を提案している。5000ノードのケースで従来の1894.3MBから13.1MBにメモリ消費量を削減できることを確認している。また、1000ノードに100ノード追加するケースで、再構築時間が11.6時間短縮(従来比18.7%)できることを示している。
- (3) 大規模な3次元画像データの高速レンダリング技術において、多数ポリゴンで構成される高解像度画像データを取り扱うためのインフラ構成と、クラスタノード間を流れる大量のデータフロー制御について、光スイッチベースの汎用PCクラスタによるSort-Lastレンダリング方式を提案している。本方式では、木構造のクラスタ構成を採用することで、光スイッチ内のスイッチング遅延問題を回避する。また、ノード間でのデータ転送に関し、フィードバック制御を行うことで、ノード間のスムーズなデータフローを実現できることを示している。

以上のように、本論文は大規模データを取り扱う分散処理システムの信頼性・処理性能向上に関する先駆的研究として、情報科学に寄与するところが大きい。よって、本論文は博士(情報科学)の学位論文として価値あるものと認める。