

氏名	大辻 弘貴		
学位の種類	博士(工学)		
学位記番号	博甲第 7699 号		
学位授与年月日	平成 28 年 3 月 25 日		
学位授与の要件	学位規則第 4 条第 1 項該当		
審査研究科	システム情報工学研究科		
学位論文題目	A Study on High-performance and Reliable Distributed Storage System (高性能かつ信頼性の高い分散ストレージシステムに関する研究)		
主査	筑波大学 教授	博士(理学)	建部 修見
副査	筑波大学 教授	工学博士	朴 泰祐
副査	東京大学 教授	博士(理学)	田浦 健次朗
副査	筑波大学 准教授	博士(工学)	追川 修一
副査	筑波大学 准教授(連携大学院)	博士(工学)	谷村 勇輔 (産業技術総合研究所)

論文の要旨

ストレージシステムが大規模化することにより、性能の向上だけではなく信頼性の向上も重要なファクタとなる。これらのファクタはトレードオフの関係にあり、両方を向上させることは難しい。本論文では、その困難を解決するため、新しい機構を提案するものである。信頼性の向上のためには、機器障害によるデータ破損に備え、冗長符号が用いられる。冗長符号は、データ転送量が増加するだけではなくパリティ計算などの演算も必要である。既存システムでは RAID コントローラがこの処理を行っており、性能低下はある程度免れていたが、ストレージシステムの大規模化に伴うストレージノード間での冗長符号では性能低下が大きな問題となっていた。本論文では、まず、ストレージノード間の RAID のためのアクティブストレージ機構を提案した。本機構は、冗長計算に伴う演算とデータ転送をストレージノードにオフロードし、パイプラインで行うものである。評価の結果、冗長符号における性能低下を抑えられることを示した。次の手法は、この冗長符号に伴う処理をネットワークスイッチにオフロードするものである。現状のスイッチに対し、そのための機構を組み込むことにより、連続書込についてはストレージノードへのオフロードよりも効率的であることを示した。ストレージシステムが更に大規模化するとストレージノード間だけではなくストレージサイト間のデータ共有が求められる。このとき、ネットワーク遅延が大きくなり遠隔のストレージシステムへのデータアクセス性能の低下が大きな問題となっていた。ネットワーク遅延の影響を少なくするためには大きいデータサイズでのアクセスが必要であるが、一方必要の無いデータ転送は無駄となり性能低下の原因となる。本論文では、この問題を解決するために、アクセスパターンに適応的なデータサイズの決定手法を提案し、静的手法の最適なケースより効率的であることを示した。

審査の要旨

【批評】

ストレージシステムの大規模化にあたり、性能の向上と信頼性の向上は重要なファクタである。信頼性の向上のためには従来から RAID 等ディスク間での冗長符号が用いられてきたが、ストレージシステムの大規模化に伴い、ディスク間ではなくストレージノード間での冗長符号が必要となってきた。ストレージノード間での冗長符号はコストが高く、これまでは書込時は複製を作成し、負荷が低い時に冗長符号化するなどの手法が用いられていた。本研究では、この問題を解決するため、ストレージノード間の冗長符号処理をクライアントからストレージノード、あるいはネットワークスイッチにオフロードする手法の提案を行っている。ストレージノードにオフロードすることにより、クライアントにおけるパリティ計算と冗長符号のデータ転送をなくすことができる。一方で、遅延時間の増加、ストレージノードの負荷の増加、ストレージノード間のネットワークトラフィックの増加などの影響を考慮する必要があり、全体としての性能向上は明らかではない。本研究では、ストレージノード間のパイプライン処理と高速な RDMA 通信によりそれらの問題の軽減を図った。3 ストレージノード、1 パリティノードのストレージノード間 RAID-5 における、8 クライアントから並列書込の評価では、冗長性のない 3 ストレージノードのストレージノード間 RAID-1 を凌ぐ性能を示しており、提案手法の有効性を示した。ネットワークスイッチへのオフロードは、ネットワークスイッチに新しくストレージノード間の冗長符号のための機構を追加する提案である。ネットワークスイッチにオフロードすることにより、連続書込においては冗長符号に伴うデータ転送を最小にすることができるようになる。本研究では、この冗長符号機能を付加したネットワークスイッチを PC サーバで実現し評価を行った。評価の結果、ほぼ理想的な性能向上を達成しており有望な手法であることを示した。また、ストレージシステムの更なる大規模化に伴う、ストレージサイト間的高速データアクセスについての提案も行った。ネットワーク遅延が大きくなるに従い、その遅延による性能低下を抑制するため、一度に取得するデータサイズを大きくする必要はあるが、不必要なデータ取得は性能低下の原因となる。本研究では、アクセスパターンを動的に検出して適応的に取得するデータサイズを変更する手法を提案した。本手法により、最適なデータサイズに固定する方法に比べ効率的であることを示した。これらの提案手法はいずれもストレージシステムの大規模化における大きな問題を解決するものとなっている。また提案手法は実機による評価により有効性が示されている。

【最終試験の結果】

平成 28 年 1 月 25 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。