

称号及び氏名	博士（工学） 兼田 泰典
学位授与の日付	平成 24 年 9 月 30 日
論文名	「大規模集約型ストレージシステムの設計技法に関する研究」
論文審査委員	主査 辻 洋
	副査 吉岡 理文
	副査 戸出 英樹
	副査 新井 利明

論文要旨

ストレージシステムに蓄積されるデータは増加し続けている。増加し続けるデータを効率よく保持し、そして効率よく管理するために、企業においては、大規模集約型ストレージシステムを用いてIT システムを構築することが一般的となっている。大規模集約型ストレージシステムとは、数百台のハードディスク装置を搭載し、計算機を接続するための数十個のネットワークインタフェースを持ち、RAID (Redundant Array of Inexpensive Disks) 技術によって書き込まれたデータを保護するためのディスクアレイ制御コントローラを有するシステムである。ディスクアレイ制御コントローラは、複数のハードディスク装置上に、計算機に提供する記憶領域であるボリュームを構成し、ネットワークインタフェースを介しこれらボリュームを計算機に提供する。大規模集約型ストレージシステムには、ストレージ専用のネットワークであるSAN (Storage Area Network) を介して、数百台規模の計算機を接続する。複数の計算機を大規模集約型ストレージシステムに接続することで、個々の計算機にハードディスク装置や専用のストレージシステムを接続しシステムを構成する場合に比べ、利用されずに無駄となる記憶容量を減らし、記憶容量の利用効率のよいIT システムを構築することができる。

一方で、大規模集約型ストレージシステムにおいては、ボリュームの配置設計や性能設計が煩雑で難しいというシステム設計上の課題がある。計算機が書き込むデータを保持するハードディスク装置は、複数のボリュームに対し記憶容量を提供することが一般的であるため、ハードディスク装置においてアクセスの競合が発生するとシーク時間や回転待ち時間に数ミリ秒から十数ミリ秒を要し、計算機におけるボリュームのアクセス性能が変動する。このボリュームの性能変動は、他の計算機からのアクセスの頻度やデータ転送量、ボリュームの配置レイアウトによって異なるため、ストレージ管理者は、競合する他の計算機からのアクセスが発生するまで把握することができない。このため、大規模集約型ストレージシステム内

にボリュームを配置設計するプロビジョニング作業を担うストレージ管理者がこのようなアクセスの競合を見越してボリュームの配置設計や性能設計等、システム設計を行うことは極めて難しい。また近年、記憶容量の利用効率のよいITシステムを構築するため、動的プロビジョニング方式（シンプロビジョニング方式）を採用する大規模集約型ストレージシステムの利用が進んでいる。動的プロビジョニング方式は、一般に、ボリュームを一定サイズのセグメントに区切り、計算機から書き込みのあったボリュームのセグメントにのみ記憶領域（ページ）を割り当てる。このページ割り当ては、ディスクアレイ制御コントローラ内の制御プログラムによって動的に行われるため、ストレージ管理者による事前のシステム設計をさらに難しいものにする。

複数のボリュームがハードディスク装置を共有していても、それぞれのボリュームの性能を独立して見積もることができる性能排他を実現できれば、ストレージ管理者による事前のボリューム性能設計を提供できる。これにより大規模集約型ストレージシステムにおいてシステム設計が難しいという課題を解決できる。また、事前のボリューム性能設計ができると、性能を資源として捉え、性能資源の計画的な分配を実現できる。さらに、性能資源の計画的な分配を発展させ、性能資源の動的な割り当て制御が実現できると、性能を必要とするボリュームに優先的に資源を割り当てる等、システム自動最適化の基盤技術として適用できる。ボリュームの性能低下等、トラブル発生時にストレージ管理者が実施するシステム再設計が難しいという課題があるが、システム自動最適化を実現できれば、この課題も解決できる。

そこで本研究は、ハードディスク装置の性能分割というコンセプトを提唱する。物理的なハードディスク装置を複数の仮想ハードディスク装置に分割し、その仮想ハードディスク装置の組み合わせによりボリュームを構成する。仮想ハードディスク装置一台一台が、物理的に一台のハードディスク装置を共有するにもかかわらず、他の仮想ハードディスク装置へのアクセスの発生によって性能の変動が起こらないように分割することができれば、物理的なハードディスク装置を占有できた場合と同様に、そのボリュームの性能を見積もることが可能となる。すなわち、ボリューム間の性能排他を実現できる。このハードディスク装置の性能分割というコンセプトを実現し、仮想ハードディスク装置の組み合わせによるシステム設計技法を実現するためには、ハードディスク装置における高い精度での性能分割を実現する必要がある。よって、本研究は、ハードディスク装置の性能分割制御方式を検討し、検討した性能分割制御の性能分割精度の検証を目的とする。また近年、大規模集約型ストレージシステムにおいて適用が進む動的プロビジョニング方式のボリュームにおける性能排他の実現方式と性能分割精度の検証、並びに、動的な性能資源の割り当て方式の提案と、動的に割り当てた性能資源の利用率の検証を目的とする。

資源の競合による性能変動の回避や、性能資源の公平な分配については、VIOS や、AVATAR/SARC, End-to-End QoS, SLEDS, Rate Window, Interposed Proportional Sharing, Deficit Round-Robin, Argon, ユーザレベルでのディスク帯域制御等が提案されている。また、システムの自動最適化については、BOAG, Self-Star Storage, ABLE, ALIS 等が提案されている。しかしながら、ハードディスク装置の性能分割によるボリュームの性能排他や、大規模集約型ストレージシステムにおけるシステム設計技法、および、動的プロビジョニング方式のボリュームの性能排他や動的な性能割り当てについては考慮されていない。

第2章では、まずはじめに、固定ボリューム構成の大規模集約型ストレージシステムにお

けるハードディスク装置でのランダムアクセス競合によるボリュームの性能変動の課題を示す。そして、ハードディスク装置の性能分割によるボリュームの性能排他というコンセプト、同コンセプトに基づくシステム設計技法を提案する。ボリュームの性能排他を実現し、同コンセプトに基づくシステム設計技法を実現するためには、ハードディスク装置における性能分割の精度が課題となる。そこで、「動的抑止時間割当方式」と「分散ボリューム配置方式」の組み合わせによるハードディスク装置における性能分割制御方式を検討し、性能変動率を評価する。

第3章では、固定ボリューム構成の大規模集約型ストレージシステムにおけるハードディスク装置でのシーケンシャルアクセス混在によるボリュームの性能変動の課題を示す。そして、シーケンシャルアクセス混在時のボリュームの性能排他実現に向けたデータのプリフェッチ方式を提案する。シーケンシャルアクセスが混在すると、シーケンシャルアクセス性能が変動し、性能分割精度が低下するだけでなく、シーケンシャルアクセス性能そのものが低くなるという課題がある。そこで、性能分割精度と高いシーケンシャルアクセス性能を両立するため、ハードディスク装置におけるシーク回数の変動が起こらない「性能分割適応プリフェッチアルゴリズム」と「ランダム分散ボリューム配置方式」による性能分割制御方式を検討し、性能変動率を評価する。第2章と第3章の評価結果から、ハードディスク装置の性能分割というコンセプト実現の見通しを得たことを示す。これにより、ボリュームの性能排他を実現でき、固定ボリューム構成の大規模集約型ストレージシステムにおいて、ストレージ管理者による事前のボリューム性能設計を実現できる見通しを得たことを示す。

第4章では、動的プロビジョニング方式を採用する大規模集約型ストレージシステムにおけるハードディスク装置でのアクセスの競合による性能変動の課題を示す。そして、動的プロビジョニング方式のボリュームにおいて性能排他を実現するため、ハードディスク装置における性能分割制御と、動的プロビジョニング方式に特徴的なページ割り当てとを連携する「性能分割連携ページ割当制御方式」を検討し、性能変動率を評価する。そして、評価の結果、動的プロビジョニング方式の大規模集約型ストレージシステムにおいても、ストレージ管理者による事前のボリューム性能設計を実現できる見通しを得たことを示す。

第5章では、動的プロビジョニング方式を採用する大規模集約型ストレージシステムにおいて、性能を資源として捉え、性能資源の動的な割り当て方式を提案する。実システムにおいては、ボリュームの必要性能は一定ではなく、ボリューム作成時に将来必要となる性能を見積もることができない場合も少なくない。このため、性能が不足し、ボリュームの性能が低下する等トラブルが発生した時に、ストレージ管理者が実施するシステム再設計が難しいという課題がある。そこで、動的プロビジョニング方式において、性能を分配や回収可能な資源として捉え、性能資源を必要なだけボリュームに割り当て、割り当てたボリュームの性能資源を高い利用率で使用する「性能資源動的割当制御方式」を検討する。そして、評価の結果、動的に割り当てたボリュームの性能資源を高い利用率で利用でき、システムの自動最適化の基盤技術として適用できる見通しを得たことを示す。

第6章では、本研究で得られた結果を総括するとともに、今後取り組むべき研究課題を整理する。

審査結果の要旨

本論文は、大規模集約型ストレージシステム的设计技法に関するもので、特に複数のボリュームがハードディスク装置を共有しているときにそれぞれのボリューム性能を独立して見積もることができるように性能排他を実現する方法について述べたものである。事前にボリューム性能設計ができると、性能を資源としてとらえ、性能資源の計画的な分配が可能となる。さらにはシステム自動最適化の基盤技術としても適用が期待できる。

本論文のオリジナリティは、「ハードディスク装置の性能分割」というコンセプトを提案し、全体性能を多少犠牲にしても、性能変動が起こらないように安定化するものである。このアイデアをもとに性能分割制御方式を論じ、その安定精度の検証するために、シミュレータを用いて、数値実験を行っており、次のような成果を得ている。

(1) 固定ボリューム構成のシステムにおけるハードディスク装置において、ランダムアクセス競合が発生した場合に起こるボリューム性能変動の課題を示し、変動を抑止するために「動的抑止時間割当方式」と「分散ボリューム配置方式」を考案し、その有効性を示している。

(2) 上記と同じ装置において、シーケンシャルアクセス混在における性能変動の課題を示し、ハードディスク装置におけるシーク回数の変動が起こらない「性能分割適応プリフェッチアルゴリズム」を考案し、その有効性を示している。

(3) ボリュームを一定サイズのセグメントに区切り、書き込み要求があったときに記憶領域を割り当てる動的プロビジョニングボリュームにおいて、性能変動の課題を示し、「性能分割連携ページ割当制御方式」を考案し、その有効性を示している。

(4) 上記と同じ動的プロビジョニングボリュームにおいて、性能を分配や回収可能な資源としてとらえ、性能資源を必要なだけボリュームに割り当て、割り当てたボリュームの性能資源を高い利用率で使用する「性能資源動的割当制御方式」を考案し、その有効性を示している。

以上の研究成果は、知能情報工学分野における大規模集約型ストレージシステム的设计技法の発展に貢献するところ大である。また、申請者が自立して研究活動を行うに必要な能力と学識を有することを証したものである。本委員会は、本論文の審査および最終試験の結果から、申請者に対して博士(工学)の学位を授与することを適当と認める。