論　文　内　容　の　要　旨

データ管理のための増加対応力に優れた分散ファイルシステム

氏　　名　荒川　淳平

情報化社会といわれる現代において，データは極めて重要な存在である．文章，画像，音楽，映像など，ありとあらゆるコンテンツがデジタル化され，電子データとして扱われている．それらのデータは私たち社会の所産であり，価値をもつ大切なものである．まさに，現代社会におけるデータは「財産」といってよいであろう．

したがって，データを管理することは財産を管理することであり，極めて重要だと言える．しかし，データの重要性が増し，私たちが日々扱うデータの種類や量が増えるにつれ，多くの問題やリスクが顕在化してきた．代表的なリスクは，装置の故障やユーザの誤った操作，コンピュータウィルスなどによるデータの損失である．

また，データ管理は個人にとどまる性質のものではない．私たちはデータを共有することで情報やアイデアを共有する．私たちはデータの漏洩を防ぎつつも，効率よく様々な人々とデータを共有しなければならない．

このように，今日，データ管理に対して私たちが要求する，あるいは要求されることは多岐にわたる．このことは，データ管理という行為自体が私たちを煩わすという事態をも引き起こしている．

この多岐にわたるデータ管理の問題に対して，本研究ではデータを 1)保管する，2)共有する，3)利用する，という 3 つの視点から問題点を整理し，それぞれについて解決策を示す．

本研究では，1)を分散ファイルシステムによって，2)を分散アクセス制御機構によって，3)をファイルシステムフレームワークによってそれぞれ解決した．そして，この分散ファイ

ルシステム，分散アクセス制御機構，ファイルシステムフレームワークの 3 つを統合したデータ管理システムによって，データ管理の抱える，多岐にわたる様々な問題を解決した．

1 番目の視点であるデータを「保管する」上で最も重要なことは，データが失われないようにすることである．ハードディスク故障によって「現在」のデータが失われることは広く知られている．また，ユーザの誤った操作などによって，必要な「過去」のデータが上書きや削除によって失われることも少なくない．そして，データを保管するシステムが障害等で利用できなくなった場合，本来であれば利用され保存されていたはずの「未来」のデータが失われたことになる．

本研究では，分散冗長化によって「現在」のデータ損失を防ぎ，ファイルシステムにバージョン管理機能を統合することで「過去」のデータ損失を防ぎ，P2P 型のアーキテクチャを採用することで単一障害点を排除し「未来」のデータの機会損失を防ぐ分散ファイルシステムを開発した．また，本研究では，ノード数に依らず一定の通信コストでデータの参照や保存を可能にすることで，データ量の増加に対応した．

ファイルシステムは多くの優れた機能を提供するが，それは決してデータ管理に必要なすべての機能ではあり得ない．データ種類や利用するデバイスの種類などが増加する限り，必要なデータ管理の機能も増加し続ける．したがって，データ管理の基盤としてとらえた場合，分散ファイルシステムは新しい機能の追加や既存機能の改良が容易でなければならない．本研究では非常にコンパクトに設計・実装することで機能数の増加に対応した．

本研究では，個々の手法としては既知の，分割と CAS によるブロック管理，追記のみのエントリ管理，Consistent Hashing を用いた分散と冗長化などを組み合わせることでコンパクトな分散ファイルシステムの構成方法を示した．また，本研究では，ローカルファイルシステムやウェブサーバの通信スタックなどを活用した実装により，実際のコード規模も小さく抑えられることを示した．

2 番目の視点である「共有する」は，企業やグループでの活動におけるデータ管理の中核的な機能である．データの共有は，共有相手のユーザに，共有したいデータに対して許可する操作が設定されることで実現される．この時，データに対する操作の可否を判断して制御するのがアクセス制御機構である．

しかし，既存の集中管理型のアクセス制御機構では，共有を行う際に必要となる，新しいユーザの登録やアクセス権限の設定に，特別なユーザ（管理者や所有者）の関与が必要である．このことは，データ量やユーザ数の増加に対して，登録・設定のための操作コストが一部のユーザに集中することを意味する．

そこで，本研究では権限証明書に基づく分散管理型のアクセス制御機構によって，ユーザによる権限の委譲を実現し，ユーザ数の増加に対応した．権限証明書はユーザ間の権限の委譲によって連鎖的に発行される．このため悪意あるユーザが見つかった場合などは，不正ユーザの権限証明書を無効化することで，不正ユーザのみならず，その不正ユーザが直接的・間接的に関わっていた権限委譲に基づくすべてのアクセスを網羅的に遮断するこ

とができる．

しかし，既存の分散アクセス制御機構では，パスワード認証が利用できないことや権限変更に副作用が伴うことで，既に広く普及している集中管理型のアクセス制御機構でのユーザ体験が活かされず，むしろ分散管理型のアクセス制御機構を利用する妨げとなっていた．そこで，本研究では新たに，公開認証情報を用いた認証，ノード秘密鍵による機構による署名，証明書の無効化と更新の区別を提案することで，既存のユーザ体験を維持したまま利用できる分散アクセス制御機構を提案した．

3 番目の視点である「利用する」は，ユーザが実際にデータ管理を行う際に生じる問題に焦点を当てる．最大の問題は「めんどくさい」である．私たちの仕事はデータ管理をすることではなく，データを利用して，または新たにデータを作り出して，仕事上の目的を達成することである．にもかかわらず，データ管理を取り巻く現状は，私たちに多くのことを要求する．定期的にバックアップを取り，個人情報は暗号化し，改訂される書類は古いデータをコピーして残しておく，などである．これでは扱うデータ量の増加や必要なデータ管理機能の増加に対応できているとは言い難い．

そこで本研究では，ファイルシステムをデータ管理システムをインターフェイスとすることで，ユーザにデータ管理を意識せずに実施させることを提案した．ユーザは普段どおりにファイルを保存するだけで暗号化などの必要なデータ管理が自動的に行われる．これにより，データ量や機能数の増加に対応し，データ管理における「めんどくさい」問題を解決することが可能となる．

しかし，従来ファイルシステムは，カーネル空間で OS に依存した形で実装されていたため，開発が困難であった．そこで本研究では，OS ごとに存在しているユーザモードファイルシステムライブラリを統合することで，ファイルシステムを，ユーザ空間で，OS に依存せず，モジュールの組合せで実装可能にした．また，ファイルシステムピボット方式を新たに提案することで，既存実装の再利用を容易にした．

以上で述べた 3 つの視点から本研究ではデータ管理の問題点を整理し，解決策を提示してきた．さらに，解決策を提示するだけではなく，本研究では常にソフトウェアとして研究内容を実装し，ユーザに利用してもらうことで多くのフィードバックを得てきた．

そして，本研究の成果である分散ファイルシステム，分散アクセス制御機構，ファイルシステムフレームワークをすべて統合したデータ管理システムは企業向けオンラインストレージサービスやストレージパッケージとしてすでに実用されている．

本研究では，データ管理の抱える数多の問題を 1)保管する，2)共有する，3)利用する，という 3 つの視点から整理し，それぞれについての問題点を解決する仕組みを提示した．1)については，「現在」・「過去」・「未来」のデータを失わず，データ量の増加にもノードの追加で対応できる機能拡張性に優れたコンパクトな分散ファイルシステムを提案した．2)については，ユーザ数やデータ量の増加に伴い増加する登録・設定コストが分散でき，柔軟で強力なアクセス制御を可能にするユーザビリティに優れた分散アクセス制御機構を提案し

た．3)については，データ管理をユーザに意識させずに，必要に応じてデータ管理の機能をモジュールとして組み合わせることで拡張可能なファイルシステムフレームワークを提案した．そして，それらを統合したデータ管理システムが実用化されたことで，その有用性を確かなものにした．