

キャッシュサーバを用いた大規模分散ファイルシステムの構築と応用

小柳 順裕[†] 山下 直人[†] 田胡 和哉[†]

東京工科大学[†]

1. はじめに

企業、公共組織、学校等、大規模組織において、その全ての構成員が利用できる、大規模な分散ファイルシステムを用いたネットワークサービスインフラを開発した。ディスクレス PC の集中管理や、分散マルチメディアリポジトリの開発を行っている。これにより、システム管理コストの軽減、セキュリティの向上、および、マルチメディアを用いたグループウェアシステムをはじめとする新しい分散アプリケーションの実現を図る。ここで用いる分散ファイルシステムを、Community Storage と名づける。

このような環境を構築することの動機の一つとして、労働形態の多様化があげられる。たとえば、政府は 2010 年までに就業人口の 2 割をテレワーカーとすることを目標に掲げている。これは就業の一部を自宅等で行うことで実現する。在宅勤務環境は、労働環境の物理的制約の軽減のためのインフラとしての観点から、デジタル社会の重要な構成要素となることが予想される。ここでは、Community Storage を用いて分散仮想オフィスシステムを構築することによって、このようなニーズに応えることをめざす。

本稿では、Community Storage とその応用、および実装について述べる。

2. システムの構造

2.1 全体システム

図 1 に、Community Storage を用いた仮想オフィスシステムの全体構造を示す。Community Storage は、1 台のサーバノードと複数のキャッシュノードから構成される。クライアントノードからは、1 台のファイルサーバと多数のキャッシュノードからなるシステム全体が、1 台の NFS ファイルサーバであるかのように見える。

キャッシュノードは、自宅やオフィスの各室に配置することを想定しており、ローカルなファイルサーバとして動作するだけでなく、Web サーバとしても運用され、クライアントノ

ードへ種々のサービスを提供する。このようなキャッシュノードの連合は、サービスグリッドとしてとらえることができる。

2.3 ディスクレス PC

Community Storage を、ネットワークブート型ディスクレス PC を集中管理するインフラとして用いる。この形式では、PC の動作やアプリケーションの実行に必要な全てのファイルを、オンデマンドにファイルサーバからダウンロードする。これは、アプリケーションの実行をクライアント側で行うので、プロセッサ負荷を分散できること、ネットワークブート可能な多くの既存の PC から利用することができること等の利点を持つ。

ディスクレス PC 運用では、仮想的に、単一のファイルシステムをすべての PC が共有する。したがって、ソフトウェアのアップデートやインストールにおいても、組織全体で一回行えば作業が完了する。このため、システムの管理コスト削減にはきわめて有効であると考えられる。また、クライアント PC が恒常的なデータを持たないために、セキュリティ向上の観点からも有

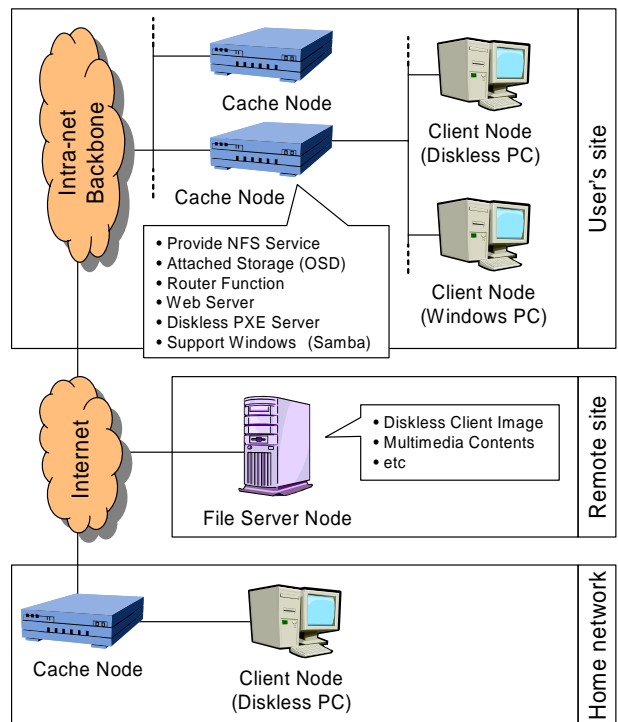


図 1 仮想オフィスシステムの全体構造

Design and Implementation of a Large-scale Distributed File System Using Cache Servers

[†] Masahiro Koyanagi, Naoto Yamashita, Kazuya Tago

• Tokyo University of Technology

効である。

2.4 分散マルチメディアリポジトリ

仮想化されたオフィスでは、業務知識を含む多くの知識は、ドキュメントよりもマルチメディアで蓄積した方が適している。Web 機構を用いて、多数の利用者の間でマルチメディアコンテンツを共有する方法を検討する。従来の、サービスを提供する Web サーバが 1 ヶ所にしかない方法では、マルチメディアによるトラフィックのためにスケーラビリティに制約があり、キャッシュ等の手段が必要になる。ここでは Web サーバを、キャッシュノードに分散して複数配置することにより、Community Storage の機能を利用して Web を用いたマルチメディアアクセスのスケーラビリティを改善する。

構築するシステムでは、共有すべきコンテンツにメタデータを付加して分散リポジトリを構成する。標準インタフェースの 1 つである JCR(Java Content Repository)^[2]を用いる。現在、JCRの分散化作業を進めており、複数のWebサーバが単一のコンテンツリポジトリを共有できる構造を構築している。

分散マルチメディアリポジトリを用いて、マルチメディアを用いたグループウェアシステムを構築する。ビューにはMozillaブラウザを拡張して利用する。JCRに保存されている情報から、コンテンツごとのスキーマを抽出し、これを、Mozillaの画面定義(XUL^[3])に変換することによって、JCRの内容を構造的に表示できるようにする。また逆に、新しいXULを定義することによって、JCRのスキーマを定義できるようにする。

3. Community Storage

ファイルサーバノードとキャッシュノードの間はNFSv4 プロトコル^[4]によって接続し、Delegation/Recall機構によってキャッシュコピーレンシを実現する。キャッシュノードに書き込まれたファイルは、比較的長期間キャッシュノード中に滞在し、ネットワークに過度の負荷をかけないことを確認しながらファイルサーバノードにライトバックされる。最終的に、ファイルサーバでファイルのバックアップを行う。

ここで構築する分散ファイルシステムは、NFS プロトコルにもとづいている。クライアントノードとサーバノードの間にキャッシュノードを挿入できるように拡張することにより、単一のサーバに接続できるクライアント数を増加させ、また、ノード間でのデータ移動を最適化することにより、マルチメディアファイル等

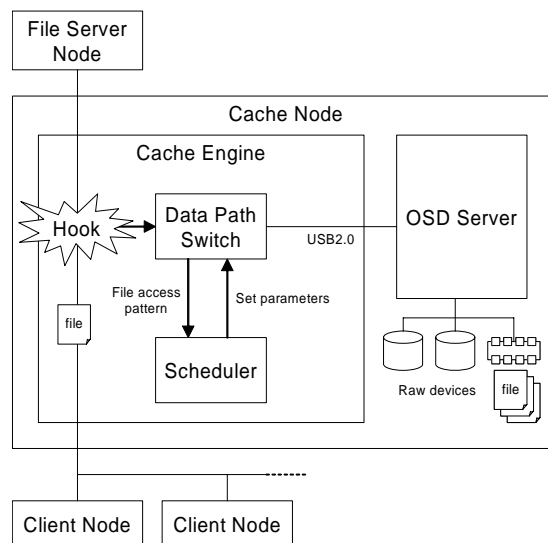


図2 キャッシュノードの論理構造

の大容量ファイルの共有を行ってもネットワークに過度の負荷がかからないようにする制御機構の実現を試みる。

図2にLinuxを用いて実装されたキャッシュノードの論理構造を示す。キャッシュ機構は、カーネルモジュールとして内実装されるDPS(Data Path Switch)と、ユーザプログラムとして実装されるSchedulerからなる。DPSは、ファイルサーバから転送されたファイルデータのキャッシュ記憶媒体への一時蓄積と、クライアントノードへのファイルエクスポートを実現する。Schedulerは、ファイルアクセスパターンの統計を解析し、自動的な性能改善を実現する。キャッシュノードは、メモリのみならず、ディスク装置や他の記憶装置を記憶媒体としてファイルキャッシュ機能を実現する。

4. 実装

現在(2006年1月)、基本機能を実装したCommunity Storage Ver.1が完成している。これを利用して、FedoraCore3を用いたディスクレスPC環境を構築した。マルチメディアを利用するグループウェアを開発した。

参考文献

- [1] 山下直人, "大規模分散ファイルシステム向きの高機能ストレージシステムの研究", 第68回全国大会
- [2] B. Callaghan, D. Robinson, R. Thurlow, Sun Microsystems, Inc., C. Beame, Hummingbird Ltd., M. Eisler, D. Noveck and Network Appliance, Inc., "RFC3530", <http://www.rfc.net/>
- [3] Mozilla.org, "XML User Interface Language", <http://www.mozilla.org/projects/xul/>
- [4] Sun Microsystems, "JSR 170: Content Repository for Java technology API", <http://www.jcp.org/en/jsr/detail?id=170>