

論文内容の要旨

博士論文題目

パラメータサーベイアプリケーションのためのメタデータを用いた
広域分散型データアクセス機構に関する研究

氏名 池部 実

(論文内容の要旨)

グリッドやクラウドなどの広域分散環境で実行した科学技術計算から出力されるデータは、インターネット上に分散するストレージへ保存される。従来、ユーザが目的のデータへアクセスする手段は、ファイル名のみが利用されてきた。しかし、広域分散環境において、他のユーザとデータ共有する際、従来のファイル名によるデータアクセス手法では、データ参照コストが大きい。データ参照コストとは、膨大なデータの中から目的のデータをユーザが発見する手間を意味する。データを生成したユーザ以外のユーザがデータアクセスする際に、データの特性やデータの持つ意味を用いることができると、ユーザは直感的にデータへアクセスすることが可能となる。本研究では、パラメータサーベイアプリケーションを対象として、データの特性や意味をメタデータで表現し、メタデータを収集・管理するための広域分散メタデータ管理システム MetaFa を提案する。MetaFa では、データを処理するワーカーノード上でアプリケーションの I/O 性能への影響を低く抑えながら、リアルタイムにメタデータを取得することができる。また、MetaFa では、メタデータスキーマを管理せずにメタデータを管理するため、様々なアプリケーションのメタデータを柔軟に管理することが可能である。

さらに、本研究ではユーザがメタデータを用いてデータ検索を行い、検索結果からデータアクセスするための PVN-FS を提案する。PVN-FS は、MetaFa が収集・管理するメタデータを用いてデータを検索する。ユーザは既存のアプリケーションを修正することなく、得られた検索結果を仮想ファイルシステム上に組み込み、データへアクセスすることが可能である。

MetaFa および PVN-FS のプロトタイプを実装し、性能評価を実施した。実験結

果から、グリッドやクラウドなどの広域分散環境において実行するパラメータサーベイアプリケーションから発生する小さなデータからメタデータを収集・管理する手法として MetaFa システムの有効性を示した。また、ユーザがメタデータを用いて柔軟にデータアクセスを行うための手法として PVN-FS の有用性を示した。本研究では、MetaFa によりグリッド環境において様々なメタデータを統一的に管理し、PVN-FS ではメタデータによる柔軟なデータアクセスを実現し、既存アプリケーションを修正することなく再利用可能とした。

本研究で提案した MetaFa は、階層型ファイルシステムなどにとらわれないメタデータを用いたデータアクセスシステムの実現に貢献すると考える。本研究では、グリッドやクラウドなどの広域分散環境におけるメタデータを用いたデータアクセスのための基盤技術を確立し、ユーザ間でデータを共有する際のデータ参照コストの削減を実現した。

(論文審査結果の要旨)

本論文は、グリッドやクラウドなどの広域分散環境でのパラメータサーベイアプリケーションの実行において、効果的なデータアクセスを実現するために、メタデータを用いたデータアクセスシステムについて議論し、広域分散メタデータ管理システム、メタデータを用いた仮想ファイルシステムを提案、評価し、システムの有効性を示した。

本論文の成果は以下に要約される。

1. 広域分散環境上で実行されるパラメータサーベイアプリケーションから生成されるデータに対して、ユーザがメタデータを用いて効果的にデータ共有・アクセスすることを目的に、広域分散メタデータ収集・管理システムに着目した。既存のメタデータ管理システムは、特定のアプリケーション向けに提供されており、異なる特性をもつアプリケーションでは利用できない問題点があった。また、既存のメタデータ収集手法では、データを処理するワーカーノードへ負荷を与え、即時的に収集することができない問題点があった。本研究では、ワーカーノードへの負荷を低く抑えながら、リアルタイムにメタデータを取得することができるメタデータ収集手法、広域分散環境で、異なる性質のアプリケーションメタデータを統一的に管理する手法を提案している。提案システムでは、既存のパラメータサーベイアプリケーションをグリッド環境上で動作させる場合の性能を満たし、広域分散環境での分散メタデータ収集・管理システムの有効性を確認している。

2. 広域に分散するデータに対して、データ検索・アクセスすることに注目し、メタデータを用いたデータ検索可能な仮想ファイルシステムの提案をしている。ユーザは既存のアプリケーションを修正することなく、得られた検索結果を仮想ファイルシステム上に組み込み、遠隔に存在するデータへアクセスすることを実現している。メタデータを用いたデータ検索を可能とし、広域でのデータアクセスを既存の広域データアクセス手法と変わらない性能を実現し、メタデータを用いた広域分散型データアクセスシステムの有効性を確認している。

以上のように、本論文はグリッド環境上で実行されるパラメータサーベイアプリケーションを対象として、広域に広がるデータに対して、メタデータを用いてユーザのデータ参照コストを削減するために 2 つの手法、広域分散メタデータ管理システム、メタデータを用いたデータアクセスシステムを提案し、評価を通してその有効性を明らかにした。学術上だけでなく、実際の環境での利用を視野に入れており、その貢献度は大きいと評価できる。よって本論文は博士(工学)の学位論文としてふさわしいものと認める。