

簡単に汎用的に利用できるグリッド計算網を構築・運用するための

ソフトウェアの開発

- 自分で始められる分散コンピューティング -

1. 背景

近年,グリッド(Grid)をキーワードとして,ネットワークを利用した様々なサービスやシステムが開発されている.これらは従来のクライアント・サーバ型のインターネットの利用から一歩進んだ概念・技術をもち,広域ネットワークにまたがって存在する各種リソース(計算資源,ストレージ資源,人的資源,サービス群,データ群など)を統合して,ユビキタスにバーチャルなコミュニティを作ることの基盤としている.その中でも分散コンピューティング,あるいはPCグリッドの分野はいち早く導入され,インターネットを通じて大規模な計算リソースを集めて計算を行うプロジェクトがいくつも立ち上がっている.しかしながら,現状,それらの基盤となるソフトウェアが商用であったり,計算リソースを提供する側がソフトウェアをインストールする手間が大変であったり,一般の個人ユーザが手軽に使えるソフトウェアは存在していない.

2. 目的

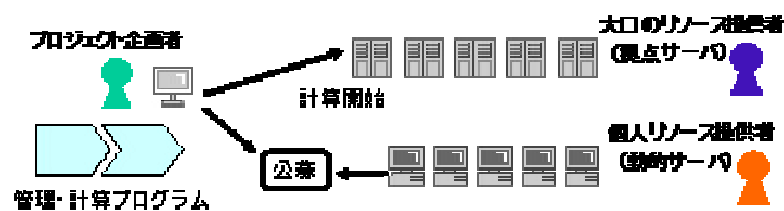
提案ソフトウェアは,大きな計算やシミュレーションのプロジェクトを開始したいプロジェクト企画者が既存のグリッド網を簡単に利用するためのミドルウェアである.そのため,複雑な設定やパラメータが極力省かれ,プロジェクト企画者が極めて簡単に分散コンピューティング・プログラムを作成し,実行できるように開発される必要がある.本プロジェクトでは,提案ソフトウェアを開発して公開することで,グリッド環境の利用のための敷居を低くし,グリッド技術の詳細を知らないプログラマが分散コンピューティング・アプリケーションを実行したり,分散コンピューティング・プロジェクトを開始したり,一般の個人ユーザがボランティアとして分散コンピューティング・プロジェクトに参加できるような環境を構築することを目指している.

3. 開発の内容

本章では,MyGMWSを構成する各モジュール群の詳細な設計および実装概要について述べ,モジュールごとに設定したテスト項目および性能評価を示す.

MyGMWSシステムは,グリッド環境を構築するために利用される最も標準的なミドルウェア Globus Toolkit をもとに構築されている.Globus Toolkit は,リソース管理,データ管理,情報サービスの基本コンポーネントをセキュリ

ティを考慮した形で提供しており、これを利用することで、グリッド・ソフトウェアを容易に開発することができる。MyGMWS では、拠点サーバで動く Agent 同士の通信の暗号化 Agent やワーカ・アプリケーションの遠隔起動に、Globus Toolkit を利用している。Globus Toolkit が利用できない環境では、MyGMWS は Secure Shell を代替として利用することができる。動的サーバの開発には、Java Web Start を利用し、Globus Toolkit の PKI の仕組みと連携が取れる形で、安全に実行ファイルを配布する仕組みを構築している。MyGMWS は、実行システムとプロジェクト企画者のための開発用ライブラリからなる。開発用ライブラリを利用して、ジョブを分配するプログラムと実際にジョブを処理するプログラムを記述することで、分散コンピューティング・プロジェクトを短時間で開始することが可能となる。



MyGMWS は、以下の手順で開発を行った。

-) プロトタイプの作成
 - プロトタイプを作成し、LAN 環境にて動作確認を行った。
 - 問題のあった箇所についてプログラムを修正した。
-) 実際のグリッド環境にて、プロトタイプの動作確認を行った。
 - 実際のグリッド環境を利用して動作確認を行った。
 - 通信ネットワークの遅延が原因となるバグを除去した。
-) プロトタイプをもとに MyGMWS 本体を開発
 - タスク情報やタスクのスケジューリングに関する部分を実装した。
 - 動的サーバのユーザ・インタフェースを実装した。
 - プロトタイプから引き継いだコア部分(通信プロトコル、ジョブの起動)について、動作確認・デバッグ作業を繰り返した。
-) GUI 部分およびデモ計算プロジェクトの開発
 - 動的サーバのユーザ・インタフェースを改良した。
 - デモ用の計算プロジェクトとして「TSP 全探索プロジェクト」を作成した。マスター側で動かすジョブ管理プログラムと、ワーカ側で動かす計算プログラムを開発した。
-) ソフトウェア全体の動作確認および品質向上にむけて全体の動作確認、デバッグ作業
 - デモ計算プロジェクトを実行し、ソフトウェアが一通り問題なく動作するかを確認を行った。
 - ソフトウェアをリリースするためのパッケージ作業の一部を行った。

分散コンピューティング・プロジェクトの開発・実行支援システムの開発ポイントとしては、以下が挙げられる。

- ・ 一般の個人ユーザがプロジェクト企画者になることができ、Globus Toolkitなどで既に構築されたグリッド環境上の大口リソースと、他の個人ユーザのリソースを取り込むことのできるシステムを開発した。
- ・ 一般の個人ユーザが Java Web Start を利用してプログラムをダウンロードし、計算を実行して、その結果を拠点サーバとに返す動的サーバプログラムを開発した。
- ・ MyGMWS の大口リソース提供者の計算機（拠点サーバ）にて Agent を遠隔起動し、自動的に Agent 同士が通信を行うシステムを開発した。
- ・ Agent 同士の通信を暗号化するために、Globus Toolkit または Secure Shell を利用できるシステムを開発した。
- ・ プロジェクト企画者が指定するパラメータによって制御されるスケジューリング機能をシステムに組み込んだ。パラメータとしては、クライアントと拠点サーバが定期的な通信を行う時間間隔、拠点サーバの最低保持タスク数、動的サーバに割り当てるタスク数の3つである。
- ・ マスター側で動かすジョブ管理プログラムは、Agent 同士の通信間隔、拠点タスク数、個人リソースに割り当てるタスク数を設定するほか、用意したタスクの入力データを設定する記述が必要である。これを作成するためのライブラリとテンプレートを開発した。
- ・ ワーク側で動かす計算プログラムは、拠点サーバから受け取ったタスクを処理し、結果を拠点サーバに返す動作を繰り返す。これを作成するためのライブラリとテンプレートを開発した。

4. 従来の技術（または機能）との相違

既存の商用ツールでは、United Device 社の Grid MP や Entropia 社の DC Grid といった商用ツールが存在するが、これらはプロジェクト企画者の経済的な負担が大きくなる傾向がある。特に、計算台数に対するライセンスの場合、多くのリソースを集めることは困難となる。

学術系のツールでは、個人リソース提供者に面倒なインストール作業を強いる傾向がある。個人リソース提供者は、コンピュータの知識をそれほど持っていない可能性があり、動的サーバのプログラムのインストールと実行は極めて簡単に行えなければならない。

5. 期待される効果

これまで、分散コンピューティングを行うアプリケーションはいくつも開発され、Web などを通じて公開されてきた。その数は決して少なくない。しかしながら、計算プロジェクト（アプリケーション）に依存せず、汎用的に使えるミドルウェアとして、オープンソースなツールが提供された例はほぼ

ないといえる。提案ソフトウェアは、分散コンピューティング分野において、有効なツールの1つとなる可能性は高いと思われる。

6. 普及（または活用）の見通し

提案ツールの有効性を理解してもらうために、普及にあたっては公開実験などを通して、成果を社会にアピールする必要がある。その際、公開実験で採用するアプリケーションが多くの人を惹きつけるものであることが大切であり、それを見て、各個人が手軽に提案ソフトウェアを扱えるようにサポート体制を整える必要がある。一方、提案ソフトウェアは多人数でリソースを共有する性質をもつため、最初の印象がよければ、比較的簡単に普及すると考える。また、公開実験以前にスケーラビリティに関する十分なテストとデバッグを行っておくことが望まれる。

7. 開発者名（所属，e-mail アドレス）

谷村勇輔(同志社大学大学院工学研究科 tanisuke@mikilab.doshisha.ac.jp)