

# 共用大規模計算機による実験用計算リソースの学内提供

宮下 夏苗

情報社会基盤研究センター

## 概要

精度の高い実験とシミュレーションは研究・教育の推進において必須の基盤となる。たとえば計算化学において複雑な物理現象，化学変化のモデリングとシミュレーションは無限の様態を示す自然現象に関する問題解決に必須の手法であり，情報，知識系各分野においても画像処理や機械学習のような大量のデータに基づく演算やリアルタイム処理，計算量の多いアルゴリズムを用いた解の抽出とアルゴリズム改良のように，今日の先端的研究において高精度な計算とそれを実現する計算リソースは必要不可欠なものといえよう。

本稿では，このように計算機を集中管理型の計算リソースとしてユーザに提供するにあたり，行っている試みについて述べる。

## 1 集中管理型の計算リソース

近年，市販 PC も安価で高性能な機器が入手可能となっており，研究室等で個別に購入したこのような機器を日常的な計算，実験に用いることも，不可能ではない。

しかし個々に購入された PC で行われる計算は，まず行える計算のサイズが PC 自身のスペックに左右される。購入後にさらに大規模な計算を行いたい要求があれば，購入しなおすか手間をかけて機器を拡張，チューニングすることになる。リソース不足が起これば計算のアルゴリズム自体を変更するなり，実験に用いるデータ量を縮小するなど，研究の本旨そのものにも関わることになりかねない。さらに，一般的な PC 機器には計算タスクの登録機能は標準では備わっておらず，多くの実験データを取るためには計算がひとつ終わるのを待ち，次の計算を投入するという，非常に手間のかかる方法をとらなくてはならない。

また，計算用に購入された機器であってもメンテナンスは必要であり，必要なアップデートやライブラリの導入，データ領域の確保とバックアップ，ソフトウェアのインストール，および故障，トラブル時のメンテナンスなど，機器を適切に利用し続けるための手間が随時必要になる。各研究室が個別に機器をメンテナンスするという事は，本来研究に使われるべき時間の一部が機器管理に割かれ，研究室と機器の個数倍，同じような時間の無駄が発生することともなりかねない。

だがそのように手間もかかり，それなりのスペックを備えた計算用機器であっても，たとえば学会終了後の閑散期，場合によっては教員や学生の転任，修了，研究テーマの変更などでリソースを空けたまま放置されることもある。

情報社会基盤研究センターは各研究室，研究者からこのような計算リソースに関わる問題を取り去り，常時充実した計算環境を提供すべく，大規模な並列計算機器群によるコンピューションサポートを展開してきた。各計算機が情報環境において最適なパフォーマンスを維持するためのベンダー各社との調整，またユーザのニーズに応じたソフトウェア・ライブラリの導入，講習会の開催等，利用と教育の双方に焦点を置いたユーザ支援を主な方針としている。

一般的なサイズより十分に大規模かつ複数のアーキテクチャからなる共用計算機環境から，ユーザはその都度希望に応じた仕様，サイズのリソースを取得することができ，ジョブの登録機能によって，時間のあるときに大量のジョブを登録しておくこともできる。また中央管理の計算機であることから，ライセンスを必要とするソフトウェアも適切なライセンス管理のもとで，各機器へのインストールを行い，提供し続けるこ

とができる。さらに、共有計算機であることから、一部のユーザが利用しない期間も他の多数のユーザが利用を続け、結果としてリソースの無駄が少なくなる。

しかし、このような共有計算機環境においても問題がある。中央管理であるために、機器の管理権限をユーザに渡すことはできず、管理権限を必要とするようなライブラリインストール、ログの閲覧などが制限される。また動作する OS が比較的なじみの薄い Linux 系であるなど、一般的なユーザにとって使いにくいと感じるだろう側面もある。

本稿ではこのような問題を克服し、計算機リソースを誰もが利用できる便利な機器としてユーザからみて「敷居を低くする」ための取組みについて述べる。

## 2 講習会

システムの安定性、また GUI のオーバーヘッドを減らす意味、また、計算用の各種ソフトウェア、ライブラリのシステム要件も考慮し、専用の計算機の OS は Linux など Unix ベースである場合が多い。しかし大多数のユーザの常用機器が WindowsOS である近年において、Linux 系の OS かつコマンドラインインタフェースでの操作を主とするような並列計算機は馴染みやすいとはいえない。

このような不自由を緩和するために、情報センターとして機器の導入ベンダー各社に依頼し、定期的にご利用者向けの講習会を開催し、講習会の調整、企画および広報を担当してきた。

しかし、従来行ってきた他社の提案を基本とした講習会のプログラムは、基本的に Unix を常用する一般的な並列計算機のユーザ、プログラマには非常に有益であるが、近年そもそも Unix に習熟しておらず、講習会の基本となるコマンドラインでの操作自体で戸惑うユーザが増えている。さらに、各種の高速化、並列化向けのライブラリを自作のプログラムに組み込むよりは、既存のフリーウェアを利用するだけで良いといったユーザが大多数を占めるようになってきた。

このような状況から従来の講習会コンテンツを見直し、Unix のコマンドライン操作に特化した講習会を別途導入する、ライブラリやオプションの詳細よりも基本的な機器の操作とサンプルの実行に重点をおき、操作方法を学ぶことに焦点をあてたわかりやすい資料準備を心がけるなど、主体的にコンテンツの提案と資料チェックを行った。

さらに、ユーザの少ないより高度な利用方法に関する講習会は回数を減らし、利用者の多いソフトウェアの利用法に焦点をあてた講習会に切り替えた。このようなソフトウェア講習会は従来の講習会の参加者に比べて多く、Matlab 講習会に至っては 50 名近い参加があったことからソフトウェア利用に対するユーザのニーズは高いことがうかがえる。

以下に 2013 年度に企画し、開催した講習会を示す。

表 1 講習会

開催時期	講習会
2013 年 6 月	並列計算機利用者オリエンテーション
2013 年 6 月	Cray XC30/MPI 初級者講習会
2013 年 6 月	SGI Altix UV1000/並列プログラミング 初級者講習会
2013 年 6 月	ScaleMP vSMP 初級者講習会
2013 年 9 月	Mathworks MATLAB 講習会
2013 年 10 月	Cray XC30/最適化プログラミング講習会
2013 年 11 月	SGI AltixUV1000/並列プログラミング講習会
2014 年 2 月	SAS ハンズオンセミナー

## 3 ソフトウェア

近年の計算リソース利用において特に顕著に見られる傾向として、商用、非商用の各種ソフトウェアの利用の増加がみられる。

研究、実験において専用に開発されたソフトウェアを利用することは、コード開発にかかる時間の大幅な短縮となり、結果の客観性の保証ともなる。情報センターもこれまで研究科から要望のあった Materials Studio など材料系のソフトウェアを中心に各種計算機へのインストール、最適化を行い、これを簡易に利用するためのスクリプトの提供、サポートを行っていた。

しかし、一般のユーザにとってそもそも計算専用機の利用自体が一般的とは言い難いため、自分の利用したいソフトウェアが並列計算機でも動作可能・高速化可能であることを想像できないため、あえて他に利用したいソフトウェアのインストールが要望として挙げられることは少ないのではないかと考えた。

そこで、2013年7月に Linux 版 Mathworks MATLAB の操作について調査を行い、PC クラスタにインストールしたうえで、利用方法を確立した。バッチシステムを利用して Matlab 計算を実行するためのサンプルスクリプトを作成し、公開を行った。並列計算機で Matlab 計算が実行できるようになったことで、これまで Matlab ユーザが手元のパソコンのみで実行していた実験、計算を大量かつ高速に実行できるようになった。

また、2013年9月には Wolfram MathMatica が GPU をサポートしていることを踏まえて、ソフトウェアの GPU クラスタ機への導入を試みた。動作検証のうえ利用方法の確立までは成功しているが、機器自体の性能が優れているとはいえないために演算自体はそこまで高速化を認められなかった。この結果を踏まえ、より高性能な機器へのソフトウェア導入と利用方法の公開を今後の課題としたい。

このほか、材料系のソフトウェアである Gaussian09 を公式にサポートしていた並列計算機 NEC SX-9 が 2014年2月をもって退役となることを踏まえ、ユーザが無理なく他の計算機に移行できるよう事前に同ソフトウェアと可視化ツールの GaussView を共有メモリ型の計算機である SGI Altix にインストールし、利用方法の確立と広報を行った。

### 3.1 インストールサポート

計算機でフリーウェアが利用できないかとの問い合わせをうけ、インストール方法の調査も行っている。2013年8月には材料系のフリーウェアである xcrysden の動作確認、2014年1,2月には同じく材料系のフリーウェアである Nwchem, NAMD を Cray XC30 向けにコンパイルする方法を調査し、依頼のあったユーザに通知した。

## 4 今後の目標

本項では、これまでに運用してきた計算機の提供形態に加え、今後より多くのユーザに利用してもらうため検討している試みについて述べる。

### 4.1 常用環境との親和性

前項で述べたように、現行のユーザは WindowsOS を常用機器としている場合が多い。また JAIST は入学当初から築かれてきた文房具としての統合情報インタフェースを全ユーザに提供しているが、このインタフェースも基本的に Windows をベースに公開されている。

これを踏まえ、Windows 上の統合開発環境である Microsoft VisualStudio でコードを開発しているユーザがシームレスに並列計算機を利用できるよう、並列計算機側の調整を行いたい。2014年8月時点では、PC クラスタのライブラリ等の環境調整を行い、VisualStudio2010 で開発したコードを PC クラスタ上でコンパイル、実行するところまで確認が取れている。一部の基幹部分のライブラリが VisualStudio のものとバージョンが合わず、関数が利用できないことが判明したため、この更新と修正を行い、Windows でコード開発、並列計算機で実験という流れを確立したい。

### 4.2 ネットワークに依存しない利用体制

また、とくに社会人学生など常時学内ネットワークに接続できないユーザも多くいる。このような不利益をなくすための方策を検討し、JAIST のユーザであれば 24 時間どこからでもストレスなく利用できる計算機環境を検討したい。可能であればどこからでも閲覧できる Web ツールを作成し、Web からの計算ジョブ

実行状況確認や、計算ジョブ投入までできるようにしたい。

## 5 まとめ

本稿では、JAIST においてユーザの研究、実験の基となる計算リソースを提供するための計算機環境について、ユーザからみてより使いやすくするための取組みと、今後の目標について述べた。

計算機リソースの運用担当者として計算機の状況と今後の課題、なによりも計算機を利用するユーザの希望につねに敏感であることがもっとも重要な課題と考える。実際にユーザと会話し、利用状況を確認し、どのような機能を実現すればよりユーザに資することができるか、常に考えていきたいと思っている。