

地球科学技術分野における組織・分野間連携促進と 計算機利用効率向上のための調査研究

(財)地球科学技術総合推進機構 堀内 一敏

1. はじめに

本調査研究では、大規模化・高精度化する地球環境の数値モデルの開発において、その組織・分野間の連携促進のためのデータや数値モデルの共有化、大量出力データの管理・効率的処理、及び計算機利用効率向上を目的として、グリッド技術を調査し、地球科学技術分野への導入を検討した。

2. スーパーコンピュータおよびその利用情報基盤

2. 1 スーパーコンピュータ開発・導入動向

米国では、地球シミュレータがそれまでの世界最高速コンピュータであった米国の記録を一気に一桁書き換えたことなどを契機に、スーパーコンピュータの国家戦略を策定する、高性能コンピューティング (High-End-Computing : HEC) 再活性化タスクフォース (HEC Revitalization Task Force) が、国家科学技術会議 (NSTC) の「ネットワーキング及び情報技術研究開発 (Networking and Information Technology Research and Development : NITRD)」プログラムの下に結成された。同タスクフォースでは、産業界で供給されているシステムが政府機関のミッションを果たすアプリケーションの性能要求を満たすには不十分とし、科学者、大学、産業界、政府機関と連携をとり、科学技術の進展を目指して今後の HEC 投資への提言をまとめた。米国政府は、これを受け、科学技術でのグローバルリーダーシップの維持のために、高性能コンピューティングを軸にした戦略を強力に進めている。このような背景の下、2011 年には、ペタスケールクラスのスーパーコンピュータとして、ASC Sequoia (アスキーセコイア) がローレンス・リバモア国立研究所 (LLNL) に導入される。また、Blue Waters が米国米国立スーパーコンピュータ応用研究所 (NCSA) で稼動する予定である。

欧州では、スーパーコンピュータ自体の開発を行っていないが、そのユーザとして、如何に効率的に活用するかを力点を置いている。現在、ペタスケールクラスのスーパーコンピュータを獲得し研究開発に活用しようとする動きが世界的に進められており、欧州でも、多くの国々が一丸となって推進している。

我が国では、理論、実験と並び、現代の科学技術の方法として確固たる地位を築きつつあるスーパーコンピューティングを更に発展させるため、長期的な国家戦略を持って取り組むべき重要技術、すなわち、国家基幹技術として「次世代スーパーコンピュータ」(以降、「次世代スパコン」) を 2012 年からの運用を目指して開発を進めている。

スーパーコンピュータの開発動向としては、純粋なベクトル計算機はなくなる方向にある。Cray 社では、ベクトルとスカラーが共存した計算機を既に販売しており、2011 年頃にはそれらがハードウェアとしてメモリーを共有する計算機を販売する予定である。スカラー機は、素子自体の性能が良いわけではなく、集積することで理論性能を達成している。IBM 社では、更なる費用対効果や省電力化を狙い、Cell 等の画像処理専用チップ等を採用した計算機を販売している。このように、アーキテクチャーが多様化する方向にある。

このようなスーパーコンピュータ開発動向を踏まえると、その性能を引き出すためには、極めて高度なアルゴリズム開発能力とプログラミング能力を持った研究者・技術者の確保と将来に向けた人材育成が重要になると考えられる。

2. 2 スーパーコンピュータ上の情報基盤の開発・導入動向

スーパーコンピュータ上の情報基盤として、主にグリッド技術を考える。グリッドという言葉は周知のとおり電力供給ネットワークのパワーグリッドに由来している。ユーザーは電力がどの発電所で作られ、どの送電線を通ってきたかを知る事はなく、必要な電力の分だけ結果としてコンセントから使う事ができる。このようにユーザー本位で資源の分配ができないかという視点から、アルゴンヌ研究所のイアン・フォスターからグリッドコンピューティング (Grid Computing) という概念が発表された。この概念は電力と同様にコンピュータの資源を利用する事ができ、さらには電力供給が現在の社会的・産業的基盤を形成しているのと同様な形でコンピューティングが社会基盤をなす時代が到来するであろうという予想に基づいている。ネットワークで計算機を接続するという発想を支えているのは、CPU の処理能力が約 18 か月で 2 倍のスピードになるというムーアの法則を超える形で、ネットワークの速度が 9 ヶ月で 2 倍のスピードになるというギルダーの法則である。従来の PC の内部バスの速度よりも速い速度のネットワークが確保されるようになった今、全てをローカルで処理するという根拠は相対的に崩れつつある。

グリッド技術は、コンピュータの高性能化を目指すものでなく、ネットワークの高速化に伴うコンピュータの利用に関する技術であり、ソフトウェアにおいて実現されるものである。グリッド技術の第一人者であるイアン・フォスターはグリッドとして認められる為の 3 つの条件を次のように提示している。

1) 分散している資源を管理できること、2) 標準プロトコル、インタフェースを利用することができること、3) 単純な寄せ集めに終わらない質的に上の次元のサービスを提供するもの。応用面から次のように分類される。①一つの巨大なコンピュータに見せかけるコンピューティング・グリッド、②遊休デスクトップ PC を有効活用するデスクトップ・グリッド、③一つの巨大なストレージのように見せかけたり仮想的に一つのデータベースに見せかけて横断的な検索を可能にするデータ・グリッド、④多次元・多地点に配置するセンサーをデータの発生源とするセンサー・グリッド、⑤仮想空間として協調作業の「場」を提供するアクセス・グリッドなどである。

グリッド技術はスーパーコンピュータをネットワークを介して利用する、米国の研究から始まっており、米国には過去のグリッド技術開発と実証システム開発の蓄積がある。その上に、さらなるグリッド技術開発、大規模な実証システム、グリッド応用開発が積極的に進められている。Globus Toolkit は、グリッドサービスを提供するソフトウェアインフラであり、それを基盤として、分散型の研究ネットワークを提供する TeraGrid の構築が全米科学財団 (NSF) により 2001 年から進められている。現在、世界最大の最も包括的で科学研究に開かれた、分散型サイバーインフラとなっている。その他、科学研究コミュニティによるグリッド形成として、気候変動研究を目的とした Earth System Grid などがある。

欧州の代表的なグリッドプロジェクトとしては、大規模データを取り扱う EU Data Grid がある。欧州原子核研究機構 (CERN) が中心となり米国、日本とも連携を取って進められている。英国以外の欧州各国でも、ドイツの Unicore、英国の e-Science プロジェクトなどがある。e-Science プロジェクトは、2000 年度を開始しており、その狙いは、科学研究を IT 技術により推進加速させることにある。

我が国では、ITBL (IT-Based Laboratory) プロジェクトが、日本原子力研究所のグリッドに類似した技術開発プロジェクト STA (Seamless Thinking Aid) の研究成果を踏まえ、文部科学省振興局の 6 年計画のプロジェクトとして 2001 年度から始められた。2001 年度から 2002 年度にかけては、産業総合研究所グリッド研究センター設立や、東工大キャンパスグリッド構築、阪大を中心としたバイオグリッドプロジェクト、東大、北陸先端大、徳島大、理化学研究所などによるバイオインフォマティクス用グリッド構築などが一斉に始まった。2003 年度には世界最高水準のグリッド構築を目指したナショナル・

リサーチグリッド・イニシアティブ（NAREGI）などの本格的なグリッド構築が進められた。

3. 地球科学技術分野における研究と計算機利用上の課題・要望

地球科学技術分野は、地球の大気、海洋、陸域、及び地球内部（地殻・マントル等）を対象に、物理・化学・生物学的な性状や機能に関して知見を集積し、それらの変動の予測を目指す学際的な学問分野である。気象・気候の予測や、二酸化炭素などの増加による地球温暖化、異常気象や気候変動がもたらす水資源・水災害への対策に資する科学的情報の提供などで社会に貢献している。地球科学技術分野は、対象が空間的、時間的に広くかつ全地球から地域までのスケールが大幅に異なる現象を扱うことや、現実世界での実験が基本的には困難であり主にスーパーコンピュータ等の計算機上のシミュレーションにたよらざるを得ないことに特徴がある。

このような地球科学技術分野の研究者にヒアリング調査を行った結果、次のような計算機利用上の課題や要望が挙げられた。①1 ランク下のテスト環境の整備、②データ保管環境の整備、③高速なネットワークの整備、④リモートログイン環境の整備、⑤可視化環境の充実、⑥ネットワークセキュリティの緩和、⑦計算科学者の技術支援、⑧利用サポート体制の充実、⑨柔軟なキューイングシステムの整備、⑩練成計算に向く計算機環境の整備、などである。

4. 地球科学技術分野における計算機利用上の課題解決の方向

3章で挙げた計算機利用上の課題・要望のいくつかは、ハードウェア環境に対する課題・要望であり、国内のスーパーコンピュータ開発に関する施策の展開とユーザ要求とのギャップの現れとも解釈できる。高額な税金を投入する以上、目に見える成果としてスーパーコンピュータ世界ランク TOP500 へのランクインを重視することは理解できるが、その後のスーパーコンピュータの有効活用を考えれば、米国の HEC 計画のように、ユーザから要求を整理し、計算機本体の重点投資から、ネットワーク、ストレージなども含めた投資のバランスを配慮し、必要な投資はすべきであろう。T2K スパコンなど既存計算機資源との連携を促進し、トータルとして、スーパーコンピュータを有効活用できるようにするための制度設計も重要である。グリッドは、制度設計を後押しするための技術である。

②のデータ保管環境の整備では、データグリッドの導入が考えられる。しかしながら、データグリッドで成功した高エネルギー分野が、再作成不可能な大規模な観測データから意味ある情報のみを抽出しハンドリング可能なサイズに圧縮した、多数のデータの世界的な共有を目指すのに対し、地球科学技術分野では、スーパーコンピュータ上の大規模シミュレーションからの大規模な出力データを、学術的発見を目的とした解析のために、研究コミュニティで共有するというものである。一連のデータのサイズ、データ圧縮の可否、共有範囲に大きな違いがある。このため、データ保管環境の整備にデータグリッド技術が適用可能かを見極めるためには、実証試験により性能面を含めた定量的な評価が必要である。

⑦の計算科学者の技術支援に関しては、ペタスケールコンピューティングにおいて米国の HEC 計画でも 100,000 プロセッサでは、ソフトウェアツール、プログラミングモデル、及びオペレーティングシステムに関わるソフトウェアの技術的不足を重視し、実質的な改善がない限り性能が期待できないとしている。地球科学分野の研究者へのヒアリングにおいても、特に次世代スパコン向けの開発に危機感を抱いているため、次世代スパコンに向けたプログラム開発・チューニング等の計算科学者の育成ならびにユーザへの技術支援体制の整備の検討が必要である。そのような検討のためにも、次世代スパコンの計算機アーキテクチャ、ネットワーク環境、プレ処理ポスト処理環境、利用条件・制限などを早期に決定し、それらの情報を公開すべきであろう。

5. 地球科学技術分野における計算機利用環境のあり方

地球科学技術分野の一層の科学的進展には、太陽活動、火山活動などの更なる素過程の組み込みや数値モデルの改良に加えて、干ばつ、豪雨、熱波、熱帯低気圧などの極端な気象現象や巨大地震の解析・予測のために数値シミュレーションの更なる高解像度化が必要である。数値シミュレーションが、このように多数の素過程を組み込み、かつ巨視から微視にいたる、より広範囲なスケールを考慮しなくてはならないことから、既に1研究者や1組織では研究ができなくなっている。すなわち、分野間や組織間の連携の必要性は既に存在すると言える。しかしながら、シミュレーションの出力データが巨大になることから、既存のネットワーク、ストレージ、可視化ソフトウェアでは、対応がますます困難となり、連携の障害となりつつある。次世代スパコンの利用においては、その傾向がさらに増大すると思われる。その解決には、ネットワークやストレージ等のハードウェアの強化が必須である。そのうえで、分野間や組織間の連携の一層の促進に、グリッド技術が適用可能な場面が、以下のとおりいくつか考えられる。

5.1 分野内連携

気候変動等の予測研究分野では、他の研究機関のシミュレーション結果と比較することで、その確からしさを確認し、差異があればモデル改善につなげる活動を行ってきた。これからの比較実験では、計算機性能の向上に伴い、シミュレーションモデルが高解像度になることから、従来の集中型のサーバだけでは支援しきれないことが危惧されている。このため、各国がディスクを拠出し合い、EARTH SYSTEM GRIDにより共有環境を構築する予定という。このように自身のデータは自身で管理するが、相互に利用し合いたい研究分野があれば、グリッド技術がその連携を支援する。

また、共同研究を進めるにあたり、シミュレーションからの大きな出力データを、ネットワークを介して自分の環境まで転送することが困難であったり、自分のところで、大きなデータを保存できないか、保存できても絵にできない環境下で研究を行っている研究者がいるとする。そのような場合、利用しているスーパーコンピュータと同じサイトか、あるいは高速なネットワークで接続された別のサイトにある、可視化サーバーを備えた環境に転送して、最終的な絵にしてから自分のところにもってこれる仕組みがあると、その研究者は、自分の研究室に居ながら結果を見ることができるようになる。このような場合、可視化サーバー、スーパーコンピュータ、ストレージサーバーなどの資源をひとつの計算機に見せかけるグリッド技術は、共同研究の推進に役立つ。

ただし、グリッド環境下におくサイト同士は、ネットワークに firewall を通さないことが運用上許可され、かつ組織を越えたアクセスが運用上許可されている必要がある。

5.2 分野横断連携

海底地震動と津波の練成や、津波や地震の揺れを建物の揺れなどの練成計算では、それぞれが異なるアーキテクチャのスーパーコンピュータの利用を想定している場合がある。次世代スパコンでは、ベクトル機とスカラー機の複合アーキテクチャを採用するので、このような練成計算を同時に実行することができる。しかしながら、1ランク下の類似したアーキテクチャーのスーパーコンピュータは見つからず、テスト環境がないということになる。

グリッド技術は、このような練成計算に対応可能で、次世代スパコンのためのテスト環境を提供でき、異分野の研究連携に役立つと考えられる。ただし、利用が複数の組織に跨る場合には、利用にあたっての組織間の合意が必要となる。