

科学技術計算とマストレージ

伊藤義彦

日本IBM・公共システム事業部・第二インダストリー・ソリューション営業部

技術革新による計算処理能力の向上と価格低下は、科学技術計算の性能向上の制限要素を演算処理ばかりでなく入出力にまで広げ、大容量のストレージ管理方法にも新しい枠組みを要求し始めている。それは、データを知的な財産として扱い財産を扱う上での利便性のプライオリティを再検討することによって構築可能であり、特注製品開発志向から商用製品利用志向へ、ハードウェア中心からソフトウェア中心へという人間の価値観の転換を促している。従来は演算処理の性能向上こそが最大の問題であったスーパーコンピューティングにおいても、人間が知的財産を管理するという観点で考えれば、もはや高速であることを望むゆえに他の要件を犠牲にして良いという環境ではない。米国ではASCIプログラムという国家計画のもとに統合的な環境を整備しており、その狙いを参照しながら、筆者のエンジニアリング・レベルの経験をまとめた。

Scientific Engineering Computing and Mass Storage

Yoshihiko Ito

Industry Solution Business Unit No.2, Public Sector, IBM Japan

The real bottleneck factor in supercomputing is being changed from computational to I/O performance by the tremendous improvement of price performance of commodity products, and this change also requires a new framework which describes how to build blocks to design efficient mass storage system. First, we should regard the data as "Intellectual Property", then newly prioritize a lot of requirements for Mass Storage System, which used to be just performance. We can naturally find there have already existed such paradigm shift such as there are more important factors than performance even in the supercomputing environment. Referring ASCI programs in the U.S., I compiled my experience as an engineer.

1. 科学技術計算とマストレージを取り巻く歴史と背景

歴史的に科学技術計算の領域は、計算機技術の発達そのものに大きな影響を与えてきた。計算処理能力の高速化および価格低下という両方の側面において、その速度は飛躍的である。半導体製造技術の技術革新の速度は一般的にムーアの法則という進歩関係が信じられており、それによれば18ヶ月で約2倍の性能向上が見込めるという。今後もこの法則が保たれていくと仮定すると、6年後には単体のマイクロプロセッサの性能が現在の1.6倍、12年後には2.56倍になる計算になる。1ギガフロップスの性能を持つ演算装置は5年後には1.6ギガフロップスになり12年後には2.56ギガフロップスになるということである。ムーアの法則による性能向上の勾配が、現実の世界に存在する計算機に対する要求を十分に満たすものであるかどうかについてはさだかではないが、少なくとも米国においてはそれは十分と見なされていないようである。例えば、米国エネルギー省(Department of Energy)のASCI(Accelerated Strategic Computing Initiative)プログラム[1][2]では、民間に委ねられた製品ベースの計算機の性能向上では十分でないために、より高い目標であるところの核兵器の貯蔵管理とい

う目標を梃子に、市場における製品の性能向上を加速するという目的が記されている。

このようなマイクロプロセッサの性能向上は、計算機を使用している世界中の科学者達にとって、より大きなシミュレーションを実現しようというモチベーションとなる。否、より大きなシミュレーションが将来の人類の幸福に貢献すると信じる科学者達の情熱が、産業界の製品ベースの計算機の性能向上を促している、と言った方が正しいのかも知れない。少なくとも、5年前では想像もできなかった大規模シミュレーションが現在の技術によって可能となったのは事実であり、おそらくは現在想像もできないような大規模なシミュレーションが5年後には民間主導の製品上で実現される期待も十分持てるであろう。しかしながら、大きなシミュレーションを実施するには演算装置が高速であるだけでは不可能であり、大容量の主記憶装置および補助記憶装置（磁気ディスク装置や磁気テープ装置など）、およびその性能向上、さらに膨大なデータを効率的に管理するソフトウェアのメカニズムを構築する技術が必須である。本文は、このような「データを人間が管理する」という社会学的な観点から、科学技術計算におけるデータ管理技術、つまりはマストレージ・システムに対する一つの思考の枠組みを、エンジニアリングの現場の経験をもとに提案するものである。

2. データ管理に関するプライオリティ

スーパーコンピューターの開発の歴史は、特注製品の開発の歴史である。とにかく高速な計算機を開発したいという科学技術計算の要求がその主な原動力であり、それが計算機産業そのもの技術革新を刺激していた。しかしながら、商用製品の性能が高速になった今日では、それをはるかに超えた高速な計算を必要とする処理は限られてきている。誰もが大規模な流体計算や超高精細画像を扱うわけではない。このことは、スーパーコンピューティングという世界が、計算機産業を刺激するような先進的な存在であった過去と異なり、超高速演算という1つのニッチ・エリア（隙間産業）を構成するようになったということを意味するのではないだろうか。ニッチ・エリアでは、市場規模が小さいために企業による継続的な製品開発投資が困難である。同時にニッチ・エリアのプラットフォームに対しては第三者のソフトウェア会社によるアプリケーション・プログラムの移行意欲が低い。計算機の製品開発においては、たとえ良いものが開発されたとしても、マーケット・シェアを獲得することができなければ、その製品の寿命は必然的に短いものとならざるをえないと言えるだろう。米国のASCIプログラムにおいても、次世代のスーパーコンピューターの開発に対して特注製品を開発する方針を取らず、多数の商用の計算機を相互結合することによって実現しているのはこのような理由によると推察される。[2]

従来、科学技術計算におけるデータ管理は杜撰な管理でも事足りた。というよりも、データ管理の必要性そのものが無かったと言えるかも知れない。データ管理の困難さよりも計算機の手速の方が先に制限となり、ほとんどの利用者は、人間の目で見ても管理可能な程度のデータしか扱ってこなかったし、また扱う必要もなかったのである。現在においては大分状況が異なる。そこで、本質的にデータとは何なのか、あるいはデータと人間との関わりを考えなおすべきであろう。データが持つ21世紀にフォーカスされるべき新しい側面を考えなければ、それを管理する方針が決定されないからである。我々がまず思い出さなければならないのは、

「データとは知的財産である」

という点である。要するに、収集されたデータとは、それに係る科学者達の所有する財産であり、一

種の価値である。ただ、一般的な金銭や固定資産と異なるのは、それが”知的”財産であるという点にあり、このことは複写することによって盗まれることがある可能性を指し示している。

財産は、それを扱う上でいくつかの利便性を要求する。データを管理する方針は、この利便性をプライオリティに基づいて決定されるべきである。データ管理においては、「管理可能性」、「可用性」、「安全性」、「性能」そして「継続性」の5項目がもっとも重要であると筆者は考える。

管理可能性とは、データが管理されていることを意味する。せつかくのデータがあっても、管理されなければアクセスすることができない。可用性とは、マストレージ・システムのダウン・タイムに関する考察である。いかに高速で管理されているシステムであっても、ダウンタイムが多ければ利便性は大きく損なわれ、マストレージ以外のシステム資源の可用性にも悪影響を及ぼす。安全性とは計算機業界で言われるセキュリティであり、これが重要であるのは、クラッカー達からの知的財産の保護である。性能は、データアクセスのトータル性能であり、これを決定するにはハードウェア、ソフトウェアを含めた様々な要素がある。継続性とは、マストレージ・システムの寿命に対する考察である。マストレージ・システムは、大量のデータを保管するために別システムへの移行が極めて困難であるが、前述の通りスーパーコンピューティングという世界はニッチ・エリアであるため、市場原理から考えてシステムの継続性は重要である。従来、これらの検討項目は等しく検討される必要があったにもかかわらず、定量的な項目である性能以外は定性的な項目として無視されがちであった。しかしながら、人間が管理するという観点から考えるならば、筆者の考えるプライオリティとしては、セキュリティ、継続性、管理可能性が高く、可用性、性能は低い。なぜならば、セキュリティ、継続性、管理可能性が損なわれた場合、マストレージ・システムそのものが無価値になる（データという財産の消失）からであり、可用性と性能に関しては利用者の利便性に係わるだけであるからである。

3. マストレージ・システムのコンポーネントと米国における活動

マストレージ・システムとはスーパーコンピューター・システムのデータ管理機能を補完する一つの独立したシステムであり、一般的にはスーパーコンピューター・システムと高速なネットワークによって接続される。マストレージ・システムという言葉は、言語的には単に大容量ストレージ・システムという意味に過ぎないが、マストレージ・システムを取り扱う学会、産業界では、テープライブラリー装置を含んだ階層型ストレージ管理システムという意味合いが強いようである。階層型ストレージ管理システムとは、属性の異なる複数種類の記憶装置（例えば磁気ディスク装置と磁気テープ装置）間において、キャッシュメカニズムを構築することである。磁気ディスク装置と磁気テープ装置との間のキャッシュメカニズムを構築するとは、頻繁にアクセスされるデータは磁気ディスク上に存在し、めったにアクセスされないデータは磁気テープ装置上に格納されるという自動化の仕組みの構築である。

当然のことながら、システムにはハードウェア、ソフトウェア双方がおおまかな構成要素として存在し、マストレージ・システムのハードウェア・コンポーネントとしては磁気ディスク装置、磁気テープ装置、磁気テープライブラリー装置、管理端末などからなる。磁気ディスク装置と磁気テープ装置の記憶装置としての最大の違いは、技術的には磁気ディスク装置はランダム・アクセス・デバイスであるのに対し、磁気テープ装置はシーケンシャル・アクセス・デバイスである点にあり、商業的には磁気テープ装置の方が磁気ディスク装置よりもバイト単価はるかに安い点にある。この2点はトレード・オフであり、システム設計においてはアクセス・パターンによってそれぞれの階層の容量が決定されな

ければならない。階層型ストレージ管理においては、データとは、実際のデータそのものとそのデータを管理するための管理データ（メタデータ）が存在する。階層構造を取るためには、データのステータスの管理が必須だからである。アクセスされたデータがどこに存在するのかがシステムが知らなければ、利用者に対するデータ・サービスはできない。ここで自明であるのは、一つデータ（ファイル）に対してアップデートがかかった場合、ファイルの大きさによらず、それなりのメタデータの更新のコストがかかるということである。このことは、マスマストレージ・システムの性能に大きく影響を及ぼす点である。

米国では、マスマストレージ技術に関し、特にソフトウェアの開発に注力したと言えるだろう。まず、ソフトウェアのコンポーネントについては、IEEE SSSWG(Storage System Standards Working Group)がIEEE Open Storage System Interconnection (OSSSI)として抽象的に定義しようと努力した。[3] 次に、ASCIプログラムのような高い科学技術計算の要求にマスマストレージ・システムの技術を合致させるためのアプローチとして、ASCI PSE(Problem Solving Environment)[1][2]のプロジェクトとしてHPSS(High Performance Storage System)を開発し、現在も維持している。HPSSはOSSSIの前名称であるIEEE Mass Storage System Reference Model Version 5を参考にして開発された。[4]

4. 結論

本文は、主に米国の技術動向のリサーチとエンジニアリング・レベルの記述であり、ある論拠に基づいた論証や実験の報告ではない。しかしながら、これからのスーパーコンピューティング環境が近い将来に必ずぶつかる問題点を産業界の一員として示唆すると共に、スーパーコンピューティングの世界では従来副次的な役割であったストレージに対するスタディを促すことを狙っている。筆者がエンジニアリング・レベルでストレージを学習しようとしても、残念ながらそれを専門的に取り扱った著作は多くは無く、現在の関連技術が進んでいる方向性や自分の思考の妥当性を検証することは邦文の文献によっては極めて困難であった。同時に、米国ではIEEEを筆頭にマスマストレージが1つのターゲット・エリアとなってスタディが進められており、実際にHPSSのようなソフトウェアも開発されている。

参考文献

- [1] "ASCI Homepage" <http://www.llnl.gov/asci/>
- [2] "ASCI 新米国スーパーコンピュータプロジェクト" 川添 良幸著、Bit 1998年12月号
- [3] "Reference Model for Open Storage System Interconnection Mass Storage System Reference Model Version 5" by IEEE Storage System Standards Working Group, 1994
http://www.ssswg.org/public_documents/MSSRM/V5-pref.html
- [4] "The Architecture of the High Performance Storage System" by Danny Teaff, Dick Watson, Bob Coyne, http://esdis-it.gsfc.nasa.gov/MSST/conf1995/A4_1.html