

研 究 解 説

統合プラットフォーム

Platform for Problem Solving Environment

小 池 秀 耀*・松 原 聖**

Hideaki KOIKE and Kiyoshi MATSUBARA

1. 背 景

現在、計算科学技術は大きな技術革新の時代にある。この技術革新の最大の特徴は、コンピュータとネットワークの発達およびこの50年間のソフトウェアや情報技術およびデータの蓄積等により、従来は不可能であった複雑で大規模な問題を解析することが可能となったことである。分散メモリ型並列コンピュータの実用化により、計算速度の飛躍的向上が可能となるとともにメモリやディスク容量も飛躍的に増大した。ENIAC以来のコンピュータの性能向上は、ピーク計算速度で約1000億倍に、記憶容量で約100億倍に向上している。コンピュータのピーク性能は、今後も5年間で10倍以上向上するものと予想され、記憶容量も同様の速度で向上するものと考えられる。また、インターネットなどのネットワーク上に蓄積されたデータやソフトウェア等を含む知識は、ここ5年間で驚異的に増大しており、誰もが容易にアクセスできるようになった。このような進歩を背景に、計算科学技術も新しい発展段階に達している。

この技術革新の中で計算科学技術が挑戦すべき中心的課題は、コンピュータの性能向上と過去の蓄積をベースにより複雑な自然現象全体を解析することに移りつつある。このための技術的核心は、ソフトウェアやデータベースなどを統合化する技術である。統合化を実現するためには、数値解法の面では多原理統合型シミュレーション技術などが必要となり、情報技術的にはタスク並列処理やネットワーク上での分散協調処理等が必要となる。なによりも重要なことは、過去50年間に渡って蓄積されてきたソフトウェアおよびデータを、ネットワーク上で統合化する技術の確立である。このような状況を背景に、GridあるいはITBLと言った分散コンピューティング環境が重要視されている。実際、複雑な現象を解析するためには、ネットワーク

上に分散した様々な計算資源（ハードウェア、ソフトウェア、データ、知識など）を有機的に活用することが必要であり、これには分散コンピューティング環境なしには実現できない。

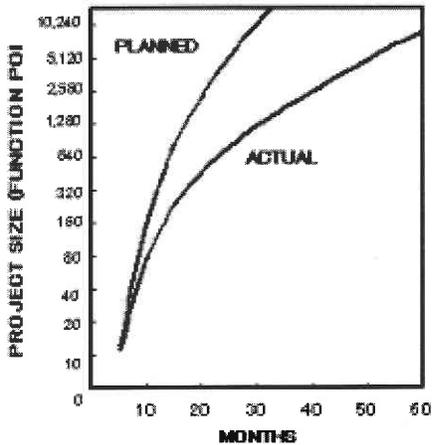
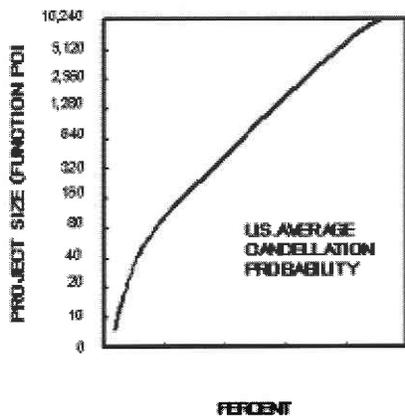
しかしながら、この技術革新を推進する上で新たな技術的困難が発生している。ソフトウェア開発がハードウェアの進歩についていけないという問題である。たとえば、タスク並列を含めた並列処理技術は全く未成熟で、ハードウェアの進歩に利用技術が追従できていないという現象が生じている。また、ネットワーク上に分散する膨大な知識から必要とする知識をいかにして抽出するかが大きな問題となっている。さらに、大規模・複雑化したソフトウェアの開発は、ますます困難になっており危機的状況に陥りつつある。表1にE.Yourdonがまとめたソフトウェア開発の難易度¹⁾を示す。最近の計算科学技術用の先端的ソフトウェアは大規模・複雑化し、現在の技術レベルは限界に近い規模および複雑さに到達している。図1は、米国の大規模ソフトウェアの開発における計画からの開発の遅れを示したものであり、大規模開発になると計画の2倍の開発期間を要している。図2は、開発プロジェクトの中止率を示したものであるが、大規模開発では半数が中止に追い込まれている。

表1 ソフトウェア開発の難易度¹⁾

ソフトウェアの規模	難易度	開発人数	開発期間
1千行以下	非常に簡単	1	数週間
1万行まで	簡単	2, 3	半年
10万行まで	やや難しい	6~10	2, 3年
1百万行まで	難しい	50~100	3~5年
1千万行まで	不可能に近い	経験者はほとんどいない	
1千万行以上	非常識	Star Wars 計画	

*東京大学生産技術研究所 情報・システム部門

** (株)富士総合研究所

図1 ソフトウェア開発と計画とのずれ²⁾図2 プロジェクトの中止率²⁾

2. PSE 研究開発の現状

計算科学技術は飛躍的發展期にあるが、ハードウェアの進歩や情報技術の蓄積を生かし技術革新を実現するためには、ソフトウェア開発の面で大きなブレークスルーを必要としている。単一のコンピュータにおける大規模ソフトウェア開発に関してもこのような困難がある。ネットワーク上の分散コンピューティング環境においては、さらに大規模で複雑な問題の解析が現在でも可能であるが、ソフトウェア開発と利用技術に関してはより大きな困難に直面することになる。したがって、ITBL や Grid 等の開発におけるデータ交換に関する基盤技術のみならず、大規模なアプリケーション・ソフトウェアの開発、利用を支援するソフトウェア・システムが必要不可欠となる。このソフトウェア・システムは PSE (Problem Solving Environment) と呼ばれている。

表2 PSE に必要とされる機能 (文献³⁾ を参考に作成)

- ① ネットワーク上に分散したコンピューティング資源 (ソフトウェア, データ) を, 統合化して, より高度で複雑なソフトウェア・システムを開発するための機能
- ② 多数のソフトウェアやデータベースを駆使した解析を, 正確に再現するための機能. すなわち, 解析手順, 使用するソフトウェア, データ, 計算結果, 中間の計算結果, 計算条件などを保存し, 解析を正確に再現できる機能
- ③ コンピュータネットワーク上に分散した, コンピューティング資源に自由にアクセスできる分散処理機能
- ④ 巨大なデータをネットワーク上で効率よく蓄積, 管理し, 必要に応じて, 必要なデータを取り出せる機能
- ⑤ ビジュアルライゼーションなどの革新的な人間インターフェイス
- ⑥ タスク並列処理ソフトウェアの開発ツール
- ⑦ 大規模シミュレーションをリアルタイムで対話的に制御するステアリング機能
- ⑧ 上記の機能をユーザーが自由に駆使できるワークベンチ

ここでは PSE を, 「プログラムの作成, 解析のセットアップ, 計算結果の蓄積と管理および後処理, 計算結果の表示のためのコンピューティング環境」と定義する. 本論文の統合プラットフォームとは, 戦略的基盤ソフトウェア用の PSE のことである. 現在の高度に発達したコンピュータネットワーク上で, 統合プラットフォームを実現するためには, 表2に示す機能を有する必要がある.

米国の高度コンピューティングに関する大規模プロジェクトでは PSE を開発の最重要課題の一つとして位置付けて研究開発を行っている. 例えば ASCI プロジェクトにおいては「ASCI シミュレーション・コードとコンピューティング・プラットフォームの最終結果は PSE にかかっている」と考えられており, ASCI プロジェクトの基盤技術として位置づけられている. また, グリッド等の分散コンピューティング環境のプロジェクトでも PSE はユーザーサイドの最上位レイヤーとして定義され, 研究開発が強力に進められている. 海外の先進的な取り組みには, つぎのような取り組みがある.

- ① Information Power Grid (IPG)⁴⁾; NASA で構築されている分散コンピューティングおよびデータマネージメント環境である. IPG ではグリッドの重要な構成要素として PSE を位置付け, PSE に関する要求仕様についても議論されている³⁾.
- ② ASCI (Advanced Strategic Computing Initiative)⁵⁾; 核実験をコンピュータ・シミュレーションで代替することを目

指す大規模なプロジェクトである。ASCIはいくつかのサブプロジェクトから構成されているが、その中の一つがASCI PSEである。ASCIではPSEの構築に力を注いでいる。

③ALICE (Advanced Large-Scale Integrated Computational Environment)⁶⁾: Argonne National Laboratoryで開発しているPSEである。ALICプロジェクトのゴールは独立に開発されたソフトウェアを統合化する上での障害を取り除くことである。

④ECCE (Extensible Computational Chemistry Environment)⁷⁾: Pacific Northwest National Laboratoryは複雑な化学システムを分子レベルで解析する統合システムMS³(Molecular Science Software Suite)を開発している。MS³はサブシステムECCEとNWChemおよびParSoftから構成されている。ECCEは化学計算用のPSEである。

3. 統合プラットフォームの開発

3.1. 目的

分散コンピューティングの大規模システムにおいては、表2に示すような機能を持つ統合プラットフォームが必要不可欠であることは共通の認識となっているが、その実現方法についてはいまだ確立されていない。CORBA等のミドルウェアの開発やXMLによるデータ交換の標準化は進んでいるが、これらはデータ交換のレベルの技術である。より重要な問題は、ネットワーク上に分散するコンピューティング資源を有機的に結合し、従来は不可能であった大規模・複雑システムを開発し、利用するためのPSEを開発することである。現在のコンピュータとネットワークの能力を活用するならば、1千万行を超える複雑で大規模なシステムを開発することは可能である。隘路はソフトウェアの開発とそのような複雑なシステムを人間が使用するための環境がないことである。現在、Gridプロジェクトではミドルウェア以下のレイヤーの開発とコンテンツの収集が中心であるが、本当の技術革新はその後にあるPSEの開発である。このような観点から考えると表2に示したPSEの機能の中でも、特に重要なのは①、②である。この機能なくして、高度に複雑で大規模なソフトウェアを開発したり、複雑な解析を自由に実行することは困難となっている。本開発の目標は①②を実現することである。

戦略的基盤ソフトウェアのような次世代の実用システムにおいては、様々なサブシステムを統合化する技術の確立が必要不可欠である。そのため、戦略的基盤ソフトウェアのPSEとして統合プラットフォームを開発する。統合プラットフォーム開発の目的は、大規模・複雑なソフトウェア統合プラットフォームの技術を確立し、具体的にインプレメントすることである。図3に分散処理における統合プラットフォームの位置付けを示す。

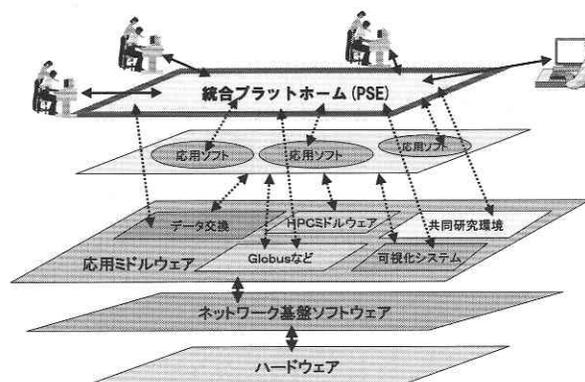


図3 統合プラットフォームの位置付け

本システムは既存のソフトウェア技術を利用している。ここで、既存のソフトウェアとは、科学技術振興調整費による総合研究「物質・材料設計のための仮想実験技術の研究」(平成7年度～11年度)⁸⁾および科学技術振興調整費による総合研究「人間系の特性を考慮した大規模・複雑システムのモデル化、解析、制御、設計に関する総合的研究」(平成10年度～14年度)において、開発されたタスクフローに関する技術である。

3.2. 特徴

PSEに対する要求仕様を実現するために、我々は「タスクフロー」を提案している⁸⁾。大規模な実用計算では、一連の処理(タスク)をある手順に従って実施することにより問題を解決する。このタスクの定義と、一連のタスクの作業手順(タスクのつながり)を定義したものがタスクフローである。タスクの定義においては、そのタスクで使用するシミュレーション・ソフトウェアやデータベースなどを定義することが可能である。当然ながら、ここで定義するソフトウェアやデータベースは、ネットワーク上に用意されていなければならない。タスクフローをまず記述し、タスクフローにしたがって解析作業を進める。この場合、中間結果の保存やタスク間のデータの受け渡しは「タスクフロー・システム」がサポートする。タスクフローの変更は自由に可能であり、タスクフローをセーブしておくことも出来る。過去のタスクフローをテンプレートとして用いることも可能である。また、タスクには解析のノウハウを記述することも可能であり、実用的な知識ベースとしてタスクフローを利用することも可能である。タスクフローにより、ネットワーク上に分散したコンピューティング資源(ソフトウェア、データ)を、統合化してより高度なソフトウェア・システムを開発する事が可能となる。すなわち、タスクフローは既存のプログラムやデータベースを部品として、より複雑で高度なシステムを構築するハイレベル言語である。既存のソフトウェアやデータベースはタスクフ

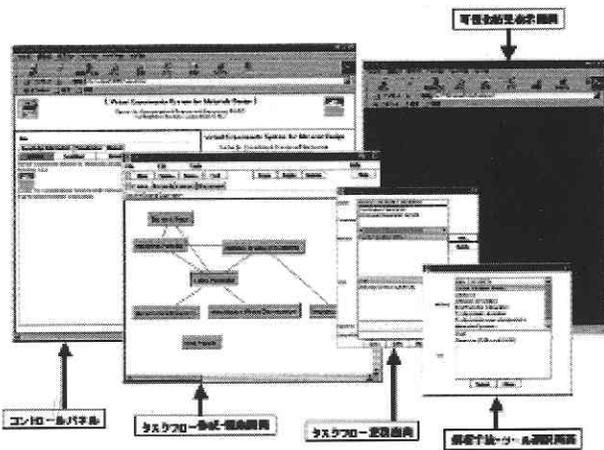


図4 タスクフロー関連の画面例⁸⁾

ローから見た場合、タスクフローが実現する大規模システムの部品である。またタスクフローにより、多数のソフトウェアやデータベースを、駆使した解析を正確に再現することが可能となる。タスクフローには解析手順、使用するソフトウェア、データ、計算結果、中間の計算結果、計算条件などが保存されている、タスクフローをセーブしておくことにより、解析を正確に再現することが可能となる。図4にタスクフロー関連の画面の例を示す。

3.3. 機能

統合プラットフォームの目的は「プログラムの作成、解析のセットアップ、計算結果の蓄積と管理および後処理、計算結果の表示のためのコンピューティング環境」を実現するソフトウェアを開発することである。PSE 実現のためには、大規模な問題を分割し、小単位で解析・設計・制御するとともに、それらの分割した小単位を再構成し統合化することが本質的に重要である。このためには、様々なソフトウェアを駆使し、複雑な手順で処理を行う必要がある。

統合プラットフォームの機能の大項目は、タスクフロー機能、タスク並列シミュレータ機能、ステアリング機能、データ管理機能の4項目からなる。それぞれの機能は、

- ①タスクフロー機能；「ネットワーク上に分散した資源を統合化して、高度で複雑なソフトウェアを開発するための機能」、「多数のソフトウェアやデータベースを駆使した解析を正確に再現するための機能」を実現する。具体的には、タスクフローの定義、タスクの定義、およびタスクフローの実行をする機能、タスクフローによる操作システム操作のログを保存および利用する機能、および、タスクフローから Web 上のリソースを利用可能とする機能を実現する。
- ②タスク並列シミュレータ機能；「ネットワーク上に分散

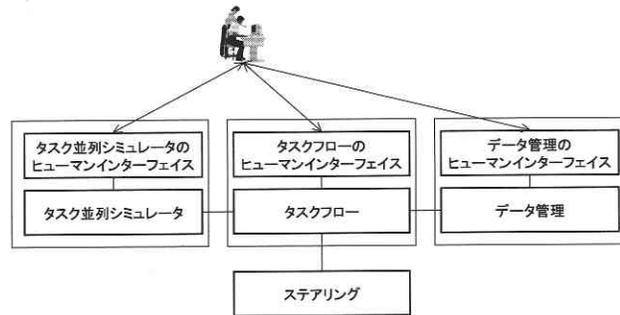


図5 統合プラットフォームのモジュール構成

した資源を統合化して、高度で複雑なソフトウェアを開発するための機能」、「タスク並列処理ソフトウェアの開発ツール」を実現する。具体的には、タスクフローで定義された操作が、並列実行の可能性をシミュレートする機能、および、タスクフロー記述されたタスクが、ネットワークを通したリソース上にて実行可能な機能を実現する。

- ③ステアリング機能；「ビジュアルイゼーションなどの革新的ヒューマンインターフェイス」および「大規模シミュレーションをリアルタイムで対話的に制御するステアリング機能」を実現する。
- ④データ管理機能；「ネットワーク上に分散した資源に自由にアクセスできる機能」および「巨大なデータをネットワーク上で効率よく蓄積、管理し、必要なデータを利用できる機能」を実現する。具体的には、タスクフローに記述されたファイルおよびその関連するファイルを管理する機能、および、タスクフローのタスクの一部として外部のデータベースを呼び出す機能を実現する。

である。これらを記述するタスクフローは言語である（言語としてのタスクフローを以降では、タスクフロー言語と呼ぶ）。タスクフロー言語を、これらの各機能共通の基盤として位置付ける。タスクフロー言語を利用して、シナリオが記述される。

3.4. タスクフロー言語

タスクフローは、XML 言語で記述され、ここではタスクフローの言語体系について示す。タスクフローは、設計作業等の作業を表現する。タスクフローは、複数のタスクから構成され、そのタスクは複数のメソッドから構成される。タスクとは、ひとつの作業のまとまりであり、そのタスクを解決するために対象物により複数の手法を利用する可能性がある。この複数の手法をメソッドであらわし、ユーザはこれらを選択する。また、このメソッドを解決するために複数のツールを選択して利用することが考えられ、これをツールであらわす。また、複数にツールは、それがまとまったタスクである可能性もあり、その場合にはツ

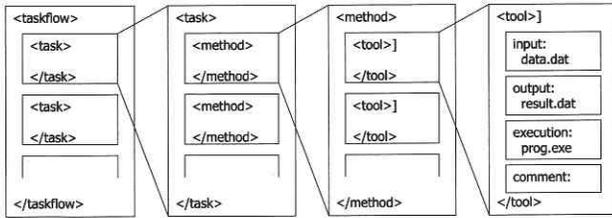


図6 タスクフロー言語の構成および内容

ルをサブタスクで表現する。図6にタスクフロー言語の構成およびその内容を示す。

4. プロトタイプの開発

4.1. 機能

現在開発したプロトタイプでは、タスクフロー基本機能のうちの一部を実現している。本プロトタイプでは、小単位を対象とする処理（これは、一般に複雑ではない）をタスクとして定義し、大規模な問題全体をタスクの連結（タスクフロー）で記述する。タスクの連結方法は多様である。タスクフローの機能は、統合プラットフォームの中核である。本プロトタイプでは、タスクの定義（分割）とタスクの組み合わせと実行順序を記述したタスクフロー（統合化）を支援する機能を実現することを主な目的としている。

具体的には、タスク定義機能では、アプリケーションやデータベース等の機能をひとつのタスクとして定義することができる。そのタスク間のリンクを定義、また作成したタスクを登録できる。サブタスク定義機能は、タスクを階層化する機能である。ここでは、現在あるタスクをサブタスクとして登録できること、および、サブタスクを持つタスクを展開して、タスクフローと同じ機能を利用できる。シナリオファイルの load と save では、タスクフロー言語を XML 形式で保存および再利用可能である。必要なファイルを保存しておけば、そこで行った作業を再利用することができる。

タスクフローの作成は、PSE 実現の方法論の確立に他ならない。本システムを用いれば、新しい方法論の確立が容易となるのみならず、最終的なタスクフローを登録することにより、高度なソフトウェアが完成する。

4.2. 操作フロー

プロトタイプでは、タスクフローを GUI 上で編集するエディタの機能を中心とする。タスクの作成、関連づけ、カット & ペーストが可能であり GUI 上にマウスでタスクを自由な位置に配置することが可能である。タスクフローは解析、設計、制御シナリオを記述したものである。タスクフローを保存することにより、シナリオ、成功事例等を蓄積、再利用が可能となる。また、図7にプロトタイプ画面例を示す。

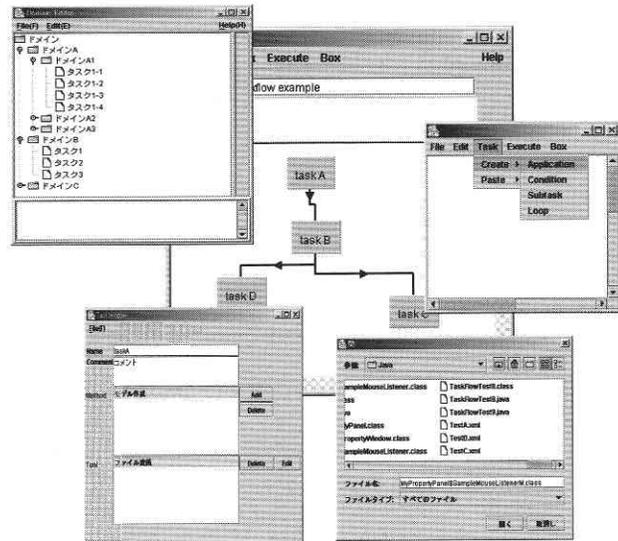


図7 プロトタイプ画面例

5. 今後の課題

ここで開発する統合プラットフォームは、産業技術、科学技術で必要とされている大規模・複雑なソフトウェアの基盤（プラットフォーム）となるものであり、その応用分野は産業、科学技術全般にわたる。現在開発済みの統合プラットフォームのプロトタイプを利用し、今後、統合プラットフォームの実証試験を実施する。実証試験では、産業機械解析システムの構築および量子計算システムの構築等を予定している。前者では、統合プラットフォームを利用して流体と構造の解析プログラムによる連成問題の解析を行うことにより実証を行う計画である。

(2003年3月24日受理)

参考文献

- 1) E.ヨードン: CASE時代の最新プロジェクト管理技術, マグロウヒル (1988).
- 2) W.W. Gibbs: Scientific American, Sept. (1994) p. 86.
- 3) W. E. Johnston *et al.*: Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid, Eighth IEEE Int. Sym. on High Performance Distributed Computing, Aug. 3-6, 1999.
- 4) <http://www.nas.nasa.gov/IPG>
- 5) <http://www.llnl.gov/ascii/PSE/>
- 6) <http://www-unix.mcs.anl.gov/alice/>
- 7) <http://www.emsl.pnl.gov:2080/capabs/mscf/software>
- 8) N. Nishikawa, C. Nagano and H. Koike: Integration of Virtual Experiment Technology for Materials Design, Computerization and Networking of Materials Databases, ATSM STP 1311, 1997