2006－HPC－107（40）
2006／8／2

Grid Datafarm における太陽地球系観測データの大規模統計解析の試み

山本 和憲[1]，村田 健史[2]，木村 映善[3]
[1]愛媛大学大学院理工学研究科
[2]愛媛大学総合情報メディアセンター
[3]愛媛大学大学院医学系研究科

太陽地球系物理分野では，データレコーダの高性能化や複数の衛星による編隊観測が行なわれるようになり，科学衛星観測で得られるデータ量が増大している．観測データの蓄積が進む一方で，CPU 性能やメモリ容量，ディスク容量を必要とする長期間の大規模な観測データの解析環境が求められている．そこで本稿では，計算グリッドとデータグリッドの特性を持つ Grid Datafarm の参照実装である Gfarm を用いて，科学衛星観測データの長期間の軌道データの解析環境の構築を行なった．実験では，逐次処理と Gfarm による 6 台のファイルシステムノードでの並列分散処理の比較を行った．その結果，1 ファイルあたり約 25 秒の処理時間を要する処理においては，実行時間を約 1/5 に短縮できた．また，並列分散処理の処理形態に着目し，科学衛星観測データ処理で，Gfarm の性能が発揮できる適用範囲について検討した．

Evaluation of Satellite Observation Data Analysis Performance with Grid Datafarm Architecture

Kazunori Yamamoto [1], Ken T. Murata [2] and Eizen Kimura [3]
[1] Graduate School of Science and Engineering, Ehime University
[2] Center for Information Technology, Ehime University
[3] School of Medicine, Ehime University

In the Solar-Terrestrial Physics (STP) field, the amount of satellite observation data has been increasing every year. More and more CPU power, memory size and disk size are required for statistical analyses of these data. To overcome this problem, we constructed a parallel and distributed data analysis environment using the Gfarm reference implementation of the Grid Datafarm architecture. Both data files and data processes are parallelized on the Gfarm with 6 file system nodes. We achieved high performance analysis as long as the data size if large enough.

1. はじめに

太陽地球系物理分野（STP: Solar-Terrestrial Physics）は，太陽活動による地球磁気圏ダイナミクスの解明を目的とした研究分野である．この分野の主な研究手法として，衛星観測と計算機シミュレーションが確立されている．特に科学衛星の直接観測（in-situ observation）によって得られるデータは，計算機シミュレーションデータとは異なり再び同じ場所と期間に観測できない．そのため，観測データの蓄積は貴重な知的財産の保存という意味がある．1992 年には太陽地

球系物理学国際共同観測計画（ISTP: International Solar-Terrestrial Physics）計画が始まり，多地点同時観測による地球磁気圏観測が行なわれてきた．衛星観測で取得されたデータは，世界各地の研究機関で分散管理される．例えば，ISTP 計画によって観測された膨大なデータの多くは，米国 NASA Goddard Space Flight Center の NSSDC データセンターで管理され，現在までに約 3.2TB のデータを公開している．また，2006 年 10 月に打ち上げが予定されている THEMIS 衛星では，1 つの衛星ミッションに 5 機の衛星が打ち上げられるため，データの蓄積量も今まで以上に増大していくことが予想される．

地球周辺の宇宙空間（geo-space）を観測した衛星観測データは，一般的な地上観測データと比べて観測点が疎であり，長期間の統計的データ解析処理による補間が重要となる．データ蓄積が進む一方，蓄積された長期間の大規模データの解析環境が求められている．しかし，統計解析処理には CPU 性能やメモリ容量，ディスク容量を必要とし，研究者が通常使用しているパーソナルコンピュータでは扱える処理量やデータサイズには限界がある．

本稿では，Grid Datafarm アーキテクチャの参照実装である Gfarm を用い，長期間の大規模衛星軌道データの並列分散処理を行う環境を構築する．また，Gfarm を用いない逐次処理の場合との比較の結果を元に，太陽地球系物理分野の科学衛星観測データ処理において，Gfarm の性能が発揮できる適用範囲について検討する．

2. 衛星観測データの解析

STP 分野の科学衛星観測データのデータフォーマットは観測計画や観測班，またはデータごとに異なっており，統一はなされていない．現在は，NASA の CDAWeb が採用している CDF（Common Data Format）形式が主流となりつつあり，今後の観測計画の多くが CDF 形式で公開されると予想される．CDF は自己記述型のデータフォーマットであり，変数やファイルの属性を記述するヘッダ部とデータ本体を記述するデータ部から構成される．

筆者らのグループではこれまでに，衛星観測データの解析システムとして STARS（Solar-Terrestrial data Analysis and Reference System）を開発してきた [1), 2)]．STARS は，データ検索および取得機能とプロット機能を提供する統合的データ解析環境である．STARS のデータ取得機能では，観測データに関する情報（データの所在地，データ期間，アクセス権限など）をメタデータベース化した STARS メタデータベースにアクセスすることにより，ユーザはファイルの分散を意識することなくデータファイルを取得することが可能である．またプロット機能では，個々のデータ読み込みプログラムをカプセル化した独自のデータクラスを持つことで，アスキー，バイナリ，画像データなど，様々なデータ形式のファイルをプロットすることができる．

観測データの解析においては，対象となる地球磁気圏構造が複雑であるため，データが観測された位置が特に重要となる．STARS では，軌道データを解析するための専用ウィンドウが用意されている．図 1 は，専用ウィンドウにおいて，GEOTAIL 衛星の 1994～1995 年の 2 年間の軌道を 2 次元断面図でプロットした結果である．軌道の色は GEOTAIL 衛星が観測したイオン密度の値をマッピングした結果を表している．

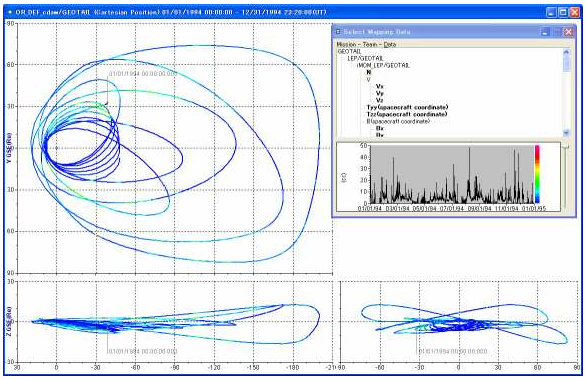

図 1　STARS の軌道専用ウィンドウ

STARS はパーソナルコンピュータ上で動作するため，CPU 性能やメモリ容量を必要とする長期間の観測データのサンプリング処理やプロットに多くの時間を

要する．また，ダウンロードされた観測データのファイルは，ローカルディスクに保存されるため，扱えるデータ量が限られる．

3. Grid Datafarm による実装

3.1 Gfarm

Gfarm[3), 4)]は Grid Datafarm の参照実装として開発されたミドルウェアである．2006 年 6 月にバージョン 1.3 が公開されている．図 2 は 3.2 で述べる本研究で構築した Gfarm 環境である．

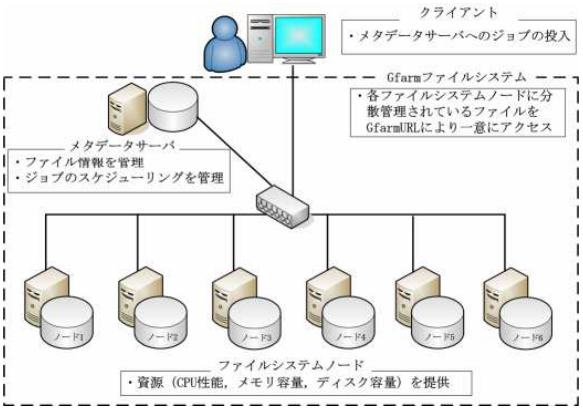

図 2　構築した Gfarm 環境

Gfarm 環境は，図 2 に示すように，メタデータサーバ，ファイルシステムノードおよびクライアントから構成される．メタデータサーバは，ファイルシステムノードに格納されているファイル情報や並列分散処理のスケジューリングを管理する．ファイルシステムノードは，プロセッサやメモリ，ディスクの資源を提供する．クライアントは，メタデータサーバへのジョブの投入を行なう．Gfarm は広域仮想ファイルシステム（Gfarm ファイルシステム）を提供しており，Gfarm ファイルシステム上のファイルを Gfarm ファイルと呼ぶ．Gfarm ファイルは GfarmURL と呼ばれるパス名で各ノードから一意にファイルアクセスすることが可能である．Gfarm ファイルは複数のファイルから構成される場合もあり，ファイルの実体は Gfarm ファイルを構成するファイル(フラグメント)として扱われ，各ファイルシステムノードのスループットディレクトリに格納される．Gfarm による並列分散処理では，各ファイルシステムノードが自ノードにあるフラグメントファイルへの処理に専念する．これにより，ローカルディスクの I/O が積極的に利用され，並列処理時におけるネットワークのトラフィックを軽減させることで，並列分散処理の性能を出している．

3.2 Gfarm 構築環境

本研究で構築した解析環境は，メタデータサーバ，6 台のファイルシステムノード，クライアントノードから成る．各ノードのスペックは表 1 の通りである．全ノード間のネットワークは 100Mbps イーサネットを用いた．

表 1　ノードのマシンスペック

	ファイルシステムノード・メタデータサーバ(6 台)	クライアント
CPU	PentiumⅢ 1GHz	Athlon 1GHz
メモリ	256MB	320MB
DISK	40GB	35GB

3.3 Gfarm ファイルシステムへの登録

本稿では，4.で述べる 2 種類の衛星の軌道データを並列分散処理する．1 つ目のデータは，GEOTAIL 衛星の軌道である．1 ファイルあたりのサイズは観測日時によって異なるが，40～80KB 程度である．2 つ目のデータは REIMEI 衛星の軌道データである．1 ファイルあたりのサイズは約 30MB である．

Gfarm では並列分散処理を行なうにあたり，Gfarm ファイルシステムに Gfarm ファイルの登録を事前に行なう．本稿では，並列分散処理を行なう際に個々の観測データファイルを指定しなくても済むように，複数のファイルを 1 つの Gfarm ファイルとして利用できる機能を用いた．この機能を用いて，長期間のデータ解析を効率よく行うために，一定の期間ごとに Gfarm ファイルを作成して登録した．登録した期間と Gfarm ファイルを構成するファイル（フラグメント）数の関係を表 2 および表 3 に示す．

ファイルシステムノードへの登録は gfreg コマンドで行なう．例えば，1994 年の 1 年間の GEOTAIL 衛

表 2　GEOTAIL 衛星データの Gfarm ファイル構成

期間	ファイル数	Gfarm ファイルサイズ[KB]
1 週間	7	315
1 ヶ月	31	1,396
半年	181	8,155
1 年	365	16,834
3 年	1,096	74,195
5 年	1,819	130,351

表 3　REIMEI 衛星データの Gfarm ファイル構成

期間	ファイル数	Gfarm ファイルサイズ[KB]
1 ヶ月	9	258,561
2 ヶ月	17	488,393
3 ヶ月	26	746,954
4 ヶ月	34	976,786
5 ヶ月	43	1,235,347

星の軌道データを登録する場合は,次のように行なう.

$ gfreg ge_or_def_1994*.cdf gfarm:ge_or_1year.cdf

これにより，1994 年の 365 個のデータファイルがファイルシステムノードに均等に分配される．また，Gfarm ファイル名 ge_or_1year.cdf を処理することによって,ユーザは Gfarm ファイルを構成するファイルを意識することなく，365 個のファイルに対して解析処理を行なうことが可能となる.

3.4　並列分散処理

Gfarm では，ファイルシステムに Gfarm ファイルを登録した場合，Gfarm ファイルを構成するフラグメント数がノード数を超える場合には，1 つのファイルシステムノードが複数のフラグメントを処理することになる．この時，並列分散処理は 2 種類の処理形態に分けられる．1 つ目は，各ファイルシステムノード内で複数のプロセスを起動させて，同時に複数のフラグメントを処理する形態である．2 つ目は，各ファイルシステムノード内で逐次的にプロセスが立ち上がり，順番にフラグメントを処理する形態である.本稿では，前者を複数プロセス並列処理と呼び，後者を単数プロセス並列処理と呼ぶことで区別する．並列分散処理を行なうには,実行プログラムが Gfarm ファイルシステム上に存在する必要がある．例えば，複数プロセス並列処理は，次のコマンドより行なわれる.

$ gfrun gfarm:〈実行プログラム名〉￥

　gfarm:〈観測データの Gfarm ファイル名〉

4. 実験

4.1　実験概要

GEOTAIL 衛星と REIMEI 衛星の軌道データについて，逐次処理と複数プロセス並列処理，単数プロセス並列処理を行なった．処理の内容は，引数として与えられた位置条件（座標値または経度・緯度）に適合する箇所をファイル走査で検索し，適合した場合には，その時刻と位置の値を標準出力するものである．例えば，図 1 において断面図から衛星が意図する領域にいる時刻を調べる場合に，X,Y,Z 座標値の最小値と最大値を入力することで，その時刻を知ることができる.

4.2　実験結果

GEOTAIL 衛星の観測データ処理実行結果を図 3 に，REIMEI 衛星の観測データ処理実行結果を図 4 に示す.GEOTAIL 衛星のデータ処理では，ファイル数によらず逐次処理の時間は 2 つの並列処理の時間に対して約 1/6 以下であった．また，複数プロセス並列処理と単数プロセス並列処理は，わずかに単数プロセス並列処理の方が短いものの，どのファイル数においても同程度の処理時間となった．なお，複数プロセス並列処理では，最大ファイル数 1,819 の場合にデータを取得できなかった.

一方，REIMEI 衛星のデータ処理では，2 つの並列処理が逐次処理に対して約 1/4 以上短いことが分かる.また，図 3 と同様に複数プロセス並列処理と単数プロセス並列処理には大きな処理時間の差はなかった．ただし，図 3 と異なり，全てのファイル数の処理におい

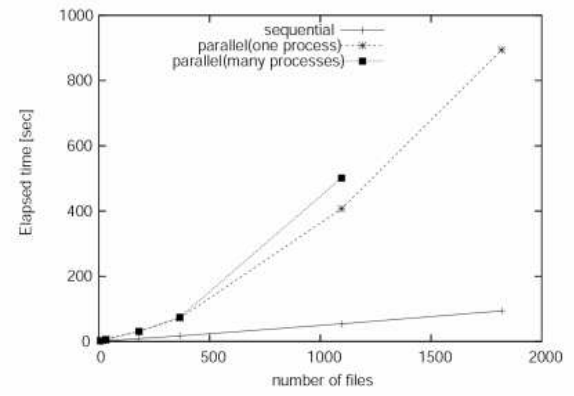

図 3　GEOTAIL 衛星の軌道データ処理におけるファイル数と実行時間の関係

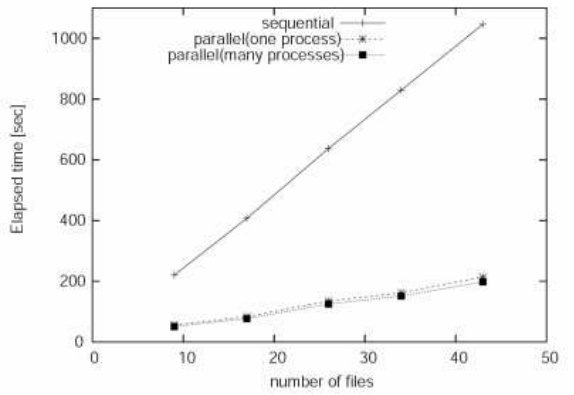

図 4　REIMEI 衛星の軌道データ処理におけるファイル数と実行時間の関係

てデータを取得することができた.

1 ファイルあたりに要する処理時間を比較すると, GEOTAIL 衛星の場合は逐次処理が約 0.05 秒, 単数プロセス並列処理は約 1 秒であった. REIMEI 衛星の場合は，逐次処理が約 24 秒，単数プロセス並列処理は約 25 秒であった.

4.3　考察

逐次処理と並列処理に分けて着目した場合，GEOTAIL 衛星の解析処理は逐次処理が有利で，REIMEI 衛星の結果は並列処理が有利であることが分かる．これは，1 ファイルに要する実行時間が，GEOTAIL 衛星と REIMEI 衛星の解析処理とでは 20 秒以上の差があるためだと考えられる．これより，1 ファイルに要する処理時間が短すぎる場合は, CPU 負荷の分散が有効に機能しないまま，処理が終了してしまっていると予想される．本稿では逐次処理と並列処理を使い分けるためのデータ取得は行なえなかったが, 1 ファイルあたり約 5 秒以上を要するデータ処理については逐次処理に比べて並列処理が有効であると予想される.

複数プロセス並列処理と単数プロセス並列処理に着目した場合，両衛星データの処理結果ともに実行時間に大きな差は見受けられない．しかし，GEOTAIL 衛星と REIMEI 衛星の結果では, 複数プロセス並列処理と単数プロセス並列処理の有効性の順番が逆転している．これは，3.4 で述べたように処理形態の違いが関係している.

単数プロセス並列処理は実効時のメモリ使用量が 1 プロセス分であるが，複数プロセス並列処理は実行時に立ち上がる複数のプロセス分のメモリ容量が必要となる．このため，本稿のようなメモリ容量が小さいスペックのノードで Gfarm 環境を構築した場合には, クライアントノードがスワップを起こしてしまう．これが，図 3 でファイル数が多くなるに従い，複数プロセス並列処理と単数プロセス並列処理で差が開く理由であると考えられる．また，図 3 のファイル数が 1,096 の場合は，複数プロセス並列処理の結果の出力数が逐次処理から得られた正しい出力数よりも少なかった．これは，スワップが影響してクライアントノードが Gfarm ファイルシステムにジョブを投入できなかったためであると予想される. さらに, ファイル数 1,819 のときのデータが取得できなかったのは，並列処理時にクライアントノードがスワップ領域を使い切ってしまい，途中からマシンが操作を受け付けなくなったためである．このように，スワップは複数プロセス並列処理における信頼性や安定性に影響している．これに対し，REIMEI 衛星の場合は，ファイル数が少ないためにスワップを起こさなかった．この場合，図 4 に見られるように複数プロセス並列処理の方が単数プロセス並列処理と比較して有効に機能する.

次に，プロセス数における 2 つの並列処理の使い分

けについて考察する．本実験では GEOTAIL 衛星の解析において，ファイル数 181 のときに両者が共に 31 秒かかっており，その後，差が開いている．クライアントのジョブ投入に使用されるメモリ容量が並列実行の処理内容や観測データのファイルサイズに依存せずに一定ならば，フラグメント数が 180 以内の解析は複数プロセス並列処理が有利であり，それ以上は単数プロセス並列処理が有利であると結論できる．

5. まとめと今後の課題

Grid Datafarm の参照実装である Gfarm を用いて衛星軌道データの並列分散処理が行なえる環境を構築し，その有効性について確認を行なった．その結果，1 ファイルあたりに要する処理時間と解析対象となるファイル数を元に逐次処理と複数プロセス並列処理および単数プロセス並列処理を使い分けることにより，効率的に長期間の観測データの解析が行なえることが分かった．具体的には，1 ファイルあたり約 1 秒の処理時間を要する解析は逐次処理し，5 秒以上の処理時間を要する解析は並列処理する．並列処理においては，1 日 1 ファイルとすると半年以内のデータは複数プロセス並列処理を行い，半年以外のデータは単数プロセス並列処理を行なうのが最適である．

今後は，逐次処理と並列処理を使い分ける具体的なデータの計測を行う．また，太陽地球系物理分野のもう一つの研究手法である計算機シミュレーションのデータ解析においても Gfarm による並列分散処理の有効性を検討する．計算機シミュレーションデータは，時空間に対して局所的な観測データとは異なり，データ生成時に時空間に制限されない分，データ量も観測データに比べて増大する．したがって，1 ファイルに要する処理時間が長くなることが予想され，Gfarm による並列分散処理の有効性を期待することができる．具体的には，図 5 に示すような 3 次元空間の中でユーザが指定した任意点の時系列グラフを描画する処理に対して Gfarm による並列分散処理を適用する．

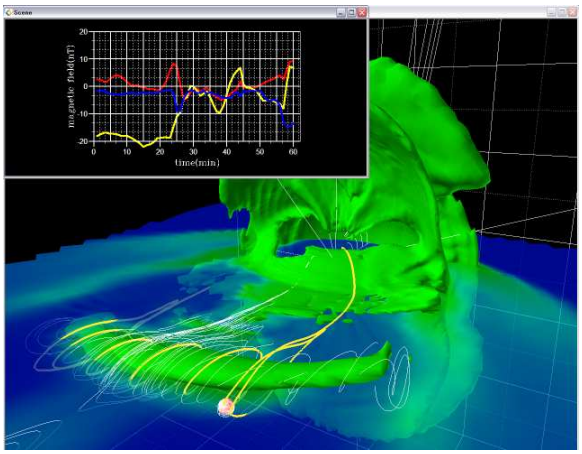

図 5　シミュレーションデータ解析（任意点の時系列グラフを描画．球体は 3 次元ポインタ．）

謝辞　本研究を遂行するにあたり貴重なご助言を頂きました Gfarm プロジェクトメンバの建部修見氏に感謝致します．本稿で用いた観測データは，宇宙航空研究開発機構・宇宙科学情報解析センターによって公開されている観測データを利用致しました．

参　考　文　献

1) 村田健史，岡田雅樹，阿部文雄，荒木徹，松本紘: 太陽地球系物理観測の分散メタデータベースの設計と評価，情報処理学会論文誌：データベース，vol.43，no.SIG12(TOD 16)，pp.115-130，Dec. 2002.

2) 村田健史: 国際太陽地球系物理観測の広域分散メタデータベース, 電子情報通信学会(B), vol.J86-B, no.7，pp.1331-1343，Jul. 2003.

3) 建部 修見，森田 洋平，松岡 聡，関口 智嗣，曽田哲之: ペタバイトスケールデータインテンシブコンピューティングのための Grid Datafarm アーキテクチャ，情報処理学会論文誌：ハイパフォーマンスコンピューティングシステム，Vol.43, No. SIG6 (HPS 5), pp. 184-195 (2002).

4) 山本 直孝，建部 修見，関口 智嗣: Grid Datafarm における天文学データ解析ツールの性能評価，情報処理学会研究報告，2003-HPC-95, pp. 185-190 (2003).