

Gfarm/Pwrake による高速可視化技術開発

山本 和憲 [1]; 村田 健史 [1]; 久保田 康文 [1]; 渡邊 英伸 [2]; 亘 慎一 [1]; 鶴川 健太郎 [3]; 村永 和哉 [3]; 国武 学 [1]; 村山 泰啓 [1]; 加藤 久雄 [4]; 建部 修見 [5]

[1] 情報通信研究機構; [2] NICT; [3] (株) セック; [4] N I C T; [5] 筑波大・シス情・CS

Development of Large-Scale Visualization System via Gfarm/Pwrake

Kazunori Yamamoto[1]; Ken T. Murata[1]; Yasubumi Kubota[1]; Hidenobu Watanabe[2]; Shinichi Watari[1]; Kentaro Ukawa[3]; Kazuya Muranaga[3]; Manabu Kunitake[1]; Yasuhiro Murayama[1]; Hisao Kato[4]; Osamu Tatebe[5]
[1] NICT; [2] NICT; [3] SEC Co.,LTD; [4] NICT; [5] Computer Science, University of Tsukuba

<http://www.nict.go.jp>

At NICT (National Institute of Information and Communications Technology) we are now developing a new research environment named "OneSpaceNet". The OneSpaceNet is a cloud-computing environment, which connects many researchers each other with high-speed network (JGN: Japan Gigabit Network). It also provides the researchers rich resources for research studies, such as super-computers, large-scale disk area, licensed applications, database and communication devices. The large-scale disk area is provided via Gfarm, which is one of the distributed file systems. One of the properties of the Gfarm is that it provides both large-scale data area and parallel processing synchronized with the data file deployment on the distributed file system. High-speed data processing is enabled to allocate data file to be read to the file system node which maintains the data file.

Computer simulation data continues being increasingly stored in many fields of geoscience, resulting in growing volumes of all kinds of data. This paper proposes a distributed data-type and/or data-intensive processing system that are provided via Gfarm/Pwrake as a solution to large-scale data processing in the context of distributed data management and data processing environments in the field of solar-terrestrial physics. The usefulness of a system composed of many file system nodes was examined using large-scale computer simulation data. In the parallel 3D visualization of computer simulation data varying in terms of data processing granularity, optimized load balancing through FIFO scheduling or pipe-line scheduling yielded parallelization efficacy.

大規模なコンピュータシミュレーションデータや多様な科学衛星・地上観測データ処理環境の実現は、次世代の宇宙天気研究へのステップアップのためには必須である。情報通信研究機構 (NICT) では、同機構において構築した科学研究向けクラウドシステムであるサイエンスクラウド (OneSpaceNet) 上において、大規模データ処理のための分散処理システムの開発に成功した。分散処理システムは、スーパーコンピュータ、分散処理・ストレージシステム (Gfarm) をクラウド上に L2 接続し、数値計算データを高速 (またはリアルタイム) に処理するシステムである。特に、Gfarm により、大規模データ格納と処理を連携することで、これまでできなかった大規模データの高速処理が可能となった。

本講演では、新しく構築した大規模可視化システムの技術について紹介する。大規模可視化システムの概要は、以下のとおりである。まず、L2 ネットワーク上のスーパーコンピュータからの大規模出力データ (たとえば、14400 時間ステップ) は、やはり L2 接続された Gfarm のファイルシステムノード (ストレージ) 上に分散して保存される。次に、各時刻ステップのデータを処理 (たとえば可視化) する。その際、各データファイルについては、データを保存しているファイルシステムノードが処理を行う。これにより、通常のデータファイルの読み込み (I/O) 時間が飛躍的に短縮できる。さらに、本システムでは、データ処理を終えたノードは、次のデータファイルの読み込みを開始する。これにより、例えば後述の MHD 磁力線追尾システムのような連続処理が必要な可視化の場合にも、高速化が可能となる。これは、データ処理を終えたノードが次の処理を行う前にデータファイルの読み込みを完了することで、ファイル I/O 時間を隠ぺいすることで実現する。

NICT では、この大規模分散ストレージ・処理システムを用いて、地球磁気圏のグローバル MHD シミュレーションデータから磁力線のフラックスを追尾するためシステム (以下、MHD 磁力線追尾システム) を構築した。MHD 磁力線追尾システムでは、次の手順により、グローバル MHD シミュレーション時空間の磁力線を追尾する。シミュレーションの全空間の全時刻ステップデータを大規模分散ストレージ・処理システム (Gfarm) に保存する。任意時刻の任意の点 (複数点でもよい) を始点群として選択する。これらの始点群と内挿により求めた各点のプラズマ速度ベクトルより、次の時刻ステップの点の位置を求める。この処理を、対象となるすべての時間ステップ分だけ繰り返し、各時刻ステップの点群データを作成する。対象となるすべての時間ステップの点群データについて、各時刻のグローバル MHD シミュレーションデータを用いて磁力線を 3 次元可視化する。各時刻ステップの磁力線可視化データをまとめて一つの時系列 3 次元可視化データを作成する。

本発表では、上記の ~ を行うための MHD 磁力線追尾システムの基盤技術について報告する。同システムは、大規模分散ストレージ・処理システム上で構築することで、実用的なシステムとして構築した。本システムの有効性を議論するため、MHD 磁力線追尾システムを一般的な大規模ストレージとワークステーション上で構築した場合と NICT サイエンスクラウド上の大規模分散ストレージ・処理システム上に構築した場合を比較する。