

TSUBAMEにおける VDI(Virtual Desktop Infrastructure)の試行

佐々木 淳, 三浦 信一, 青木 尊之

東京工業大学 学術国際情報センター

sasaki@gsic.titech.ac.jp

概要: スパコンの高性能化とともに計算規模が増大し、出力される結果データは TB を超えることも珍しくない。データ処理や可視化などのポスト処理を行うため、ローカルのワークステーション等に結果データを転送すると、データ転送時間、ローカルストレージのコストなどが膨大になり、大きな問題となっている。そこで、TSUBAMEのストレージに直接アクセスできる VDI(Virtual Desktop Infrastructure)を導入し、計算の結果データを移動することなく、ポストプロセッシング処理までスパコンで行い、その表示すべき画像データだけをローカルのクライアントに転送するシステムを試行的に導入した。CAE アプリケーションを例にとり、スパコンのクラウド的な利用が可能になり、遠隔からのスパコン利用の効率化を示す。

1 はじめに

TSUBAME は東京工業大学学術国際情報センターが 2006 年より運用しているスパコンであり、2010 年 10 月にリプレースされた TSUBAME2.0 は、同年 11 月の TOP500 ランキング 4 位、理論性能 2.4PFlops を有する日本初のペタスケールのスパコンである。2 基の Intel Xeon CPU と NVIDIA GPU アクセラレータ Tesla を 3 基搭載した Thin ノードが 1442 台を有し、高速な Fat-Tree 型の InfiniBand ネットワークにより接続されている。2013 年 9 月に全ての GPU を Kepler コアの Tesla K20X に換装する事で TSUBAME2.5 へアップグレードし、倍精度演算の理論性能 5.7PFlops、単精度演算での理論性能は 17.1PFlops へと大きな性能向上を果たした。TSUBAME の利用は、基本的にインターネットを介して SSH (セキュアシェル) 接続し、ジョブを投入する型のバッチ処理である。

スパコンの高性能化とともに計算規模が増大し、出力される結果データは数 TB を超えることも珍しくない。それらは、計算ノードから接続される高速で並列な読書きが可能な Lustre ファイルシステムで構成された 3.6PB のストレージに出力される。1 課題グループ当たり 30TB まで利用可能であるが、一番大きな問題は可視化を伴うポスト処理である。たとえ Gigabit Ethernet の学内 LAN で接続された研究室のワークステーションへのデータ転送であっても、ゆうに 1 日以上かかる。また、研究室

にデータを転送するにしても、30TB のデータを保存するストレージを確保することは容易なことではない。TSUBAME の大規模計算により出力されたデータに対して、TSUBAME のストレージから移動することなしに可視化を含んだポストプロセッシング処理を行いたい要望が急増している。

2 TSUBAME の Lustre に直結する VDI システム

利用者がリモート環境から TSUBAME をより効率的に利用できる環境として、図 1 に示す Virtual Desktop Infrastructure (以下「VDI」という) システムを試行的に導入した。VDI 環境として、Intel Xeon E5-2660v2 2 基と NVIDIA Grid K2 1 基が搭載されたホスティングサーバ 3 台を用意し、本環境上に VMWare Horizon View^[1] および Citrix Xen Desktop^[2]の環境を構築した。本システムで最大 20 クライアントまでの仮想デスクトップ環境を提供する。本環境で VDI 上から InfiniBand 等の高速ネットワークで大容量ストレージに直接アクセスできることが重要である。プリプロセッシング処理も、計算も、ポストプロセッシング処理も TSUBAME 上でクラウド的に行うことのできる環境である。TSUBAME の Lustre ストレージに直結する VDI システム (以下「TSUBAME 直結 VDI システム」という) が今後の TSUBAME ユーザの標準的な利用環境になって行くと思っている。

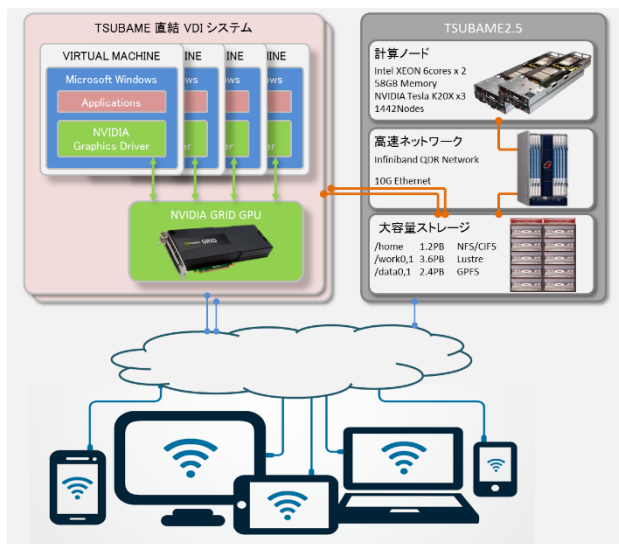


図 1 TSUBAME での VDI の構成

VDI はサーバ上にデスクトップ環境を仮想化して集約したものである。利用者はクライアント機となるノートパソコンもしくはシンクライアント端末からサーバ上の仮想マシンに接続し、仮想マシンのデスクトップ画面を呼び出して操作する。サーバ上での仮想化されたデスクトップ環境でアプリケーションを動作させるため、クライアント機にアプリケーションを導入する必要が無い。また、クライアント機の性能にアプリケーションの動作は依存しないため、非力なタブレット端末やスマートフォン等からの利用も可能である。

VDI はクライアント機に表示する画面をサーバ側で作成するが、3D レンダリングをとまなう表示等は CPU で処理するには重過ぎる。このためサーバ側に搭載した GPU により、仮想デスクトップ上に表示する画面を描画するソリューションの一つが NVIDIA Grid[®]である。オフィス系のアプリケーションのみならず、Open GL および DirectX での 3D レンダリング等の負荷の高い描画処理も可能となっている。また GPU も仮想化され、例えば NVIDIA Grid K2 1 基で最大 8 つの仮想デスクトップの描画を担うことが可能である。

3 CAE アプリケーションにおけるスパコン利用

3.1 VDI 環境での CAE アプリケーション

多くの CAE アプリケーションベンダーがクラウドコンピューティングに対応可能なライセンスを提供している。大規模計算の需要も増え、スパコンにおいてもクラウド的な利用サービスが求め

られつつある。CAE アプリケーションの利用においても、超高速計算性能、超大容量ストレージというスパコンの能力を最大限に享受できるクラウド環境の提供が必要である。従来、スパコンはその全ての計算資源を要してようやく解が得られるような問題に用いられていたが、最近はクラウドコンピューティングのリソースとしての期待も大きい。TSUBAME では、Intel x86 アーキテクチャに対応した数多くの CAE アプリケーションが稼働しており、最近では GPU アクセラレータに対応した CAE アプリケーションも増加している。CAE での大規模な解析においては多くの計算資源を同時に必要とすることから、スパコンのクラウド的な利用に対するニーズが高まってきた。

殆どの CAE アプリケーションはプリプロセッシング・ポストプロセッシングの機能を有しており、モデルの作成や種々のパラメータの設定をプリプロセッシング機能で行い、ソルバ機能にて計算を実行し、その結果をポストプロセッシング機能により解析する。ワークステーションのようなスタンドアロンシステムでは、一連のワークフローは同一の環境で実施されるが、インタラクティブな操作を伴うプリプロセッシング処理およびポストプロセッシング処理をスパコン上で行うことは出来ない。このため計算をスパコン上で行う際には、図 2 で示すように、ローカルの計算機環境上のプリプロセッシング機能で作成したモデルファイルや計算条件の設定ファイルをスパコンにアップロードして計算し、その計算結果のファイルをローカルの計算機環境にダウンロードしてポストプロセッシング機能により確認するという、非常に効率の悪いワークフローとなる。

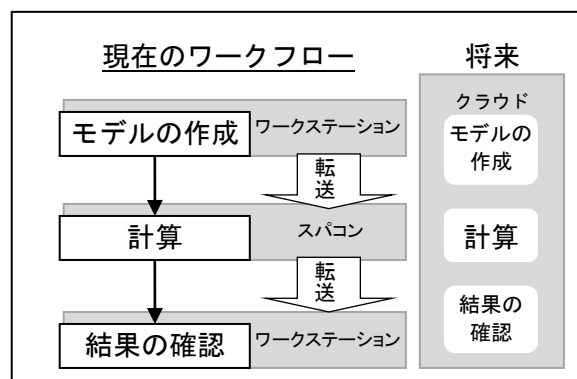


図 2 スパコンでの CAE アプリ利用のワークフロー

一部のスパコンを保有する機関においては、

計算結果の可視化のための大画面専用システムを導入しているが、CAEアプリケーションのプリ・ポストプロセッシング機能の利用は想定されていない。限られた可視化ソフトウェアやOpenGLライブラリを用いて描画するアプリケーションのみが対応しており、リモートでの可視化は非常に限られた環境でしか利用できない。また可視化システムがスパコンの大規模ストレージに直接アクセスできない場合もあり、その場合は計算結果のファイルを可視化システムへ移動させる必要がある。

3.2 ユーザ事例での評価

TSUBAME 産業利用トライアルユースでの、三ツ星ベルト株式会社「産業用ゴムベルトの有限要素法による構造解析」の事例を示す。

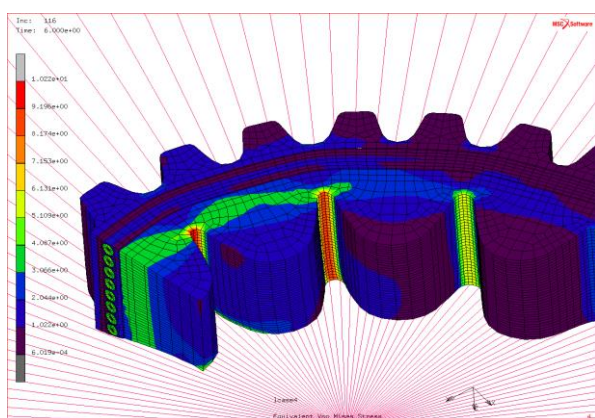


図3 産業用ゴムベルトの解析例

図3は要素数1,177,600、節点数1,185,602の産業用ゴムベルトの有限要素モデルをMSCソフトウェア社の構造解析ソフトMarcにて解析した結果である。駆動プリー内でのベルトMises応力分布を表示している。TSUBAMEにて48ノード48GPU並列で解析した際には、並列化およびGPUを使わない場合に比べ最大27倍の速度向上効果を確認している^[4]。この計算結果のファイルは24GBあり、TSUBAMEからインターネットを介して企業のワークステーションにダウンロードするには平均8MB/secで約50分を要する。

図4は、TSUBAME直結VDIシステムでMarcのポストプログラムであるMentatを操作しているスクリーンショットである。計算結果はTSUBAMEの高速大容量ストレージを構成するLustreファイルシステム上に格納されており、ポスト機能Mentatが計算結果ファイルをLustreファイルシステムから直接読みこんでいる。

NVIDIA Gridによりモデルの移動や回転等の操作もスムーズに行える。

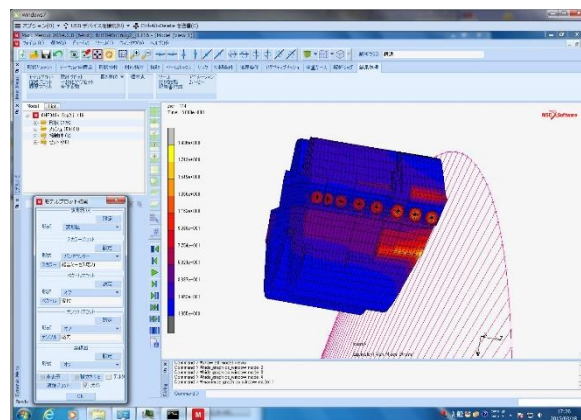


図4 TSUBAME直結VDIシステムでのポスト処理

また図5はタブレットにてTSUBAME直結VDIシステムで動作するMarcのポストプログラムMentatを操作している写真である。クライアントデバイスの能力によらず大規模計算の結果を同じように表示確認できる。

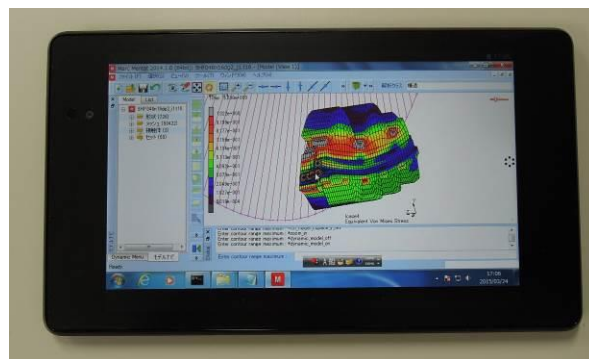


図5 タブレットでの利用例

4 まとめ

スパコンをクラウド的に利用する際の環境としてVDIとNVIDIA Gridで実現したTSUBAME直結VDIシステムを構築した。

TSUBAME直結VDIシステムの利用により、プリプロセッシング機能で作成するモデルのデータ、および計算結果のデータを含む全てのデータを一元的にクラウド環境に置くことができ、スパコンを効率的に利用できるとともに、データ移動にともなうセキュリティリスクを軽減できるというメリットもある。

謝辞

東京工業大学学術国際情報センターで進めるTSUBAME産業利用は、文部科学省の先端研究

基盤共用・プラットフォーム形成事業の補助を受け実施している。ここに記して、謝意を表す。

参考文献

- [1] VMWare, Horizon with NVIDIA GRID vGPU,
<http://www.vmware.com/jp/products/horizon-view/features/vGPU.html>
- [2] Citrix, XenDesktop,
<https://www.citrix.co.jp/products/xendesktop/overview.html>
- [3] NVIDIA, VIRTUAL GPU TECHNOLOGY,
<http://www.nvidia.com/object/virtual-gpus.html>
- [4] 徳田明彦, 三ツ星ベルト株式会社: 産業用ゴムベルトの有限要素法による構造解析, 平成25年度 TSUBMAE トライアルユース成果報告書