

巻頭言

情報基盤センター長
田浦健次朗

この巻頭言では基盤センターのスーパーコンピュータと商用クラウドの違いについて整理したいと思います。

情報基盤センターは長年、先進的なスーパーコンピュータの運用を通じて科学技術の発展に貢献してきました。現在は Intel メニーコアクラスタ(Oakforest-PACS)、Intel Xeon クラスタ(Reedbush-U)、NVIDIA P100 GPU クラスタ(Reedbush-H/L)をサービス中で、本年から新しいシステム(Oakbridge-II (仮称))も稼働します。それらのシステムは研究課題審査や、課金に基づくものなど様々な形態で全国の産学の利用に供しています。

一方現在は商用クラウドの普及・発展が著しく、AI、特に深層学習などで高速な計算機が必要という時に真っ先に使うのは商用クラウドの利用という時代になっています。なお商用クラウドにお金を払うのに比べ、基盤センターの利用負担金は、単純に計算能力あたりの値段という観点からは格安です(例えば GPU を 4 つ搭載した計算ノード 1 時間あたりの利用料は、基盤センターの場合約 35 円、某有名商用クラウドでは 492 円です。なお、機種は前者のほうが新しい)。しかも後者の場合ノードが立ち上がっている時間に対する課金であるのに対し、前者の場合は実際に計算(ジョブ)が走っている時間に対する課金です。したがって、基盤センターが幅広い分野の研究者・利用者をサポートすることは社会的にも重要な使命で、特に近年 HPC に対する需要の著しい機械学習分野のユーザに浸透させることの意義は大きいと考えます。

浸透率という意味で言うと、科学技術計算を中心とした、いわゆる HPC 分野に比べて、それ以外の分野への浸透率は低いのが現状です。HPC 科学技術計算の分野では、ごく小規模な段階を除いては、基盤センターのような、大規模で共有されたスーパーコンピュータを利用することはもはや当然になっています。一方、AI、機械学習となると、商用クラウドの利用、それだと割高になるくらい処理量が多い場合は研究室単位で数 GPU 搭載の計算機を買う、ということがまだまだ一般的と言えます。機械学習(正確には、深層学習)で行っている計算は、密行列中心の数値最適化で、スーパーコンピュータがもってこいのワークロードであるにもかかわらず、です。

なぜでしょうか？複合的な理由があると思いますが、計算機の内部構成以外の部分の違いが大きいと思います。現在の基盤センタースパコンの欠点を述べます。

- (1) 利用の障壁・利用開始までの書類・審査ラグが大きい。一方商用クラウドであれば使いたいと思ったら直ちにサインアップして使い始めることができます。基盤センタースパコンには利用申し込み、審査という手順が必要です。
- (2) 環境設定の自由度がない。一方商用クラウドは各ユーザが管理者権限を持ったノードを自由に設定できます。
- (3) 全ユーザが同じオペレーティング環境を共有するため、機密データに関する懸念が生じやすい。一方商用クラウドで使うノードのオペレーティング環境は自分だけがログインできるノードで、ネットワーク的にも他のクラウドユーザのノードと隔離されています。

機械学習応用などを想定すると(2)は重要で、同分野では千差万別の機械学習フレームワークやソフトウェアスタックが使われ、かつそれらと GPU のドライバや通信ライブラリとの相性の問題が生じやすいため、無事アプリケーションが動くまでかなりの労力を要するケースが多いです。そのような事情と、(1)の利用開始の書類の手間やラグの大きさが組み合わさって、浸透がなかなか進まないという事情があると思われます。

まず(1)については、これについては今後利用申込をウェブで完結するようにする、少量のリソースを企業からの利用者を含め簡易な審査でトライアル利用できるようにするなどの仕組みを試行的に導入していく予定です。これによって潜在的ユーザはまず自分のプログラムが動くか、高速に動くか、などをまず確かめることができます。

(2)を実現するためには、クラウドの基盤技術である仮想マシンやコンテナなどの技術を導入することが第一歩です。基盤センタースパコンでも一部のマシン(Reedbush)ですでにコンテナ技術が利用可能になっています。しかしたとえコンテナの中であっても管理者権限でソフトウェアを動かすことはセキュリティ上の問題がある(そのノードから、そのマシン内の他のノードへの特権的なアクセスを許すことになる)ため不可能です。

これを可能にするのが(3)で述べたネットワークによる隔離です。これは仮想化環境で使われている技術ですが、これとスーパーコンピュータの生命線である 1 マイクロ秒レベルの低遅延通信、台数を増やしたときの高高速化(スケーラビリティ)、ファイル IO などを両立させなくてはなりません。これらの実現に必要な技術開発を進めつつ、申請の障壁を減らすなどの改善はどんどんやっていくことで、情報基盤センターは今後も先進的な高性能計算環境をユーザに提供し続けます。今後とも皆さまのご支援を賜りますよう、お願い申し上げます。