

仮想 GPU 技術を用いた単一システム上での仮想 GPU クラスタの構築

2110507 濱野 晃暉 成見研究室

1 はじめに

近年, GPU を用いて汎用的な計算を行う GPGPU を利用したアプリケーションが増加している. 深層学習や, 画像処理などを利用するアプリケーションが増え, GPU による計算の需要が高まっている. GPU を用いて大規模な計算を行うには MPI や CUDA といったプログラミングモデルを用いて GPU クラスタを利用する必要があるが, その学習コストは高く, 教育用としてそれらを手軽に試せる環境が必要とされる.

また, GPU 仮想化の技術が進化しているが, 仮想クラスタにおいて GPU 仮想化を利用した例は少ない. GPU 仮想化を利用した仮想 GPU クラスタを構築することで, GPU クラスタを用意する際のコストや管理の手間を省くことができるのに加え, より柔軟に構成を変更することができるようになる. 加えて, GPU クラスタを用いたプログラムの性能や特性の検証を行う際にも, 仮想 GPU クラスタを利用することで, 容易に検証を行うことができる.

本研究では, GPU 仮想化を利用した仮想 GPU クラスタを構築し, その性能や特性の評価, UI による仮想 GPU クラスタ構築システムの実装を目的とする.

2 GPU クラスタ

GPGPU とは, 汎用的な計算を GPU で行う技術である. 本来 GPU はグラフィックス処理のための専用プロセッサであるが, 高い並列性を持つため汎用的な計算にも利用されるようになった. 科学技術計算や機械学習などで広く利用されており, 従来の CPU での計算と比較して高い性能や電力効率が期待できる. GPGPU で計算を行うためには, フレームワークや API によらずほとんどの場合プログラムを作り直す必要があり, その際にはハードウェア特有の制約を考慮する必要がある. GPU クラスタは, 複数の GPU を搭載したコンピュータ (ノード) をネットワーク経由で接続し, 並列計算や大規模データ処理を行うための計算基盤である (図 2). 1 つのコンピュータでは扱いきれなかったり, 長い時間を必要としたりするような処理を高速に行うことができる.

3 仮想化技術

仮想化とは, 物理的なリソースを抽象化し, 複数の論理的なリソースとして扱う技術である. 1 台のマシンを複数に分割したり, 複数台のマシンを 1 台として扱ったりする

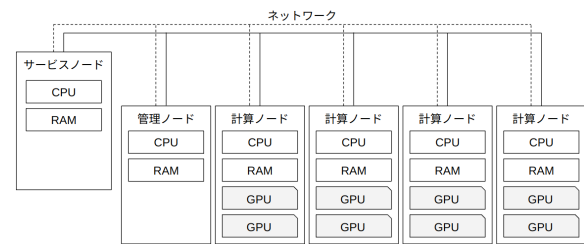


図1 GPU クラスタの構成の例

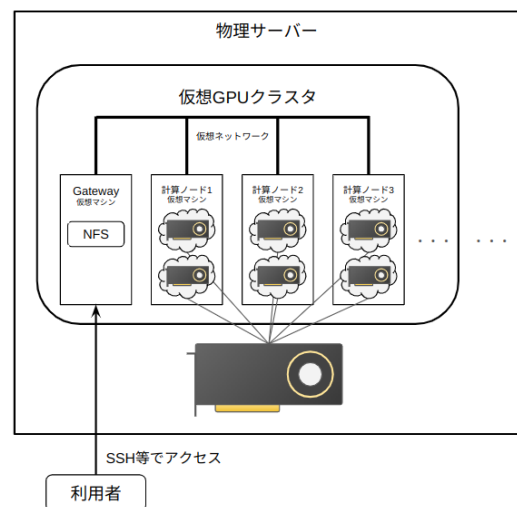


図2 仮想 GPU クラスタの概要図

ことが可能になる. 仮想化を利用することで, 物理的なリソースの利用率の最適化やセキュリティの向上, リソースの隔離などの利点を得ることができる. 近年, GPU においても仮想化の利用が可能になっており, 1 つの GPU デバイスを仮想的に分割して仮想的な GPU デバイスに分割することができる. 例として, NVIDIA 社が提供している NVIDIA Virtual GPU[1] が挙げられる.

4 仮想 GPU クラスタ管理システム

本研究では, Linux KVM/QEMU/libvirt と NVIDIA vGPU を用いて仮想 GPU クラスタの構築を行った. 仮想 GPU クラスタについて, その概要を図 2 に示す. 物理的には 1 台である GPU をラウンドロビンスケジューリングで時間分割し, 仮想 GPU デバイスを仮想マシンに割り当て, それらを仮想ネットワークで接続することで仮想 GPU クラスタを構築する.

また, 仮想 GPU クラスタの管理を WebUI を通じて行うことのできるシステムを開発した. 図 3 にそのポータル画面を示す. このポータルを通じて, 仮想 GPU クラス

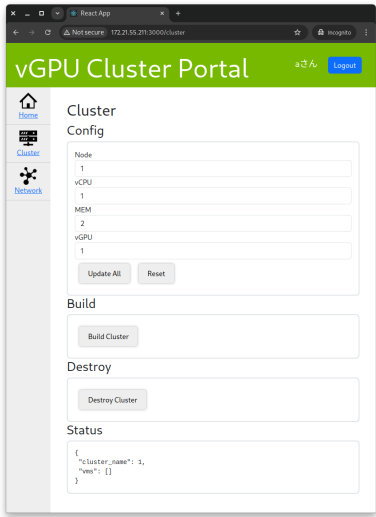


図 3 仮想 GPU クラスタ管理システムのポータル画面

ターのノード数, CPU 数, メモリ容量, 仮想 GPU 数の変更に加え, 仮想 GPU クラスタの起動, 停止, 削除などの操作を行うことができる。

5 評価

本研究では, 以下の 3 点の項目で評価を行った。

- 1. NVIDIA vGPU による GPU 仮想化の性能評価
- 2. 仮想 GPU クラスタの性能評価
- 3. 仮想 GPU クラスタ管理システムの評価

1 の評価では, 仮想 GPU クラスタは物理 GPU クラスタと比較して演算性能やメモリ帯域において一定のオーバーヘッドが確認された。演算性能 (FP32) では 14%, メモリ帯域 (GPU 内) では 10% のオーバーヘッドが確認された。メモリ帯域 (CPU-GPU 間) では PCIe 帯域の制約により, vGPU 数が増えると 1vGPU あたりの通信帯域が低下する傾向が見られた。

2 の評価では, ネットワーク帯域を適切に制御することで, 物理クラスタに近いスケーラビリティを示すことが確認された。図 4 のように仮想 GPU クラスタのネットワーク帯域を物理 GPU クラスタと同様の速度に制限すると物理 GPU クラスタと似た性能向上グラフとなった。図 5 は計算がボトルネックとなるようなアプリケーションの場合の性能であり, スケーリング傾向に差異が見られた。

3 の評価では, 開発した Web GUI による仮想 GPU クラスタ管理システムはクラスタ構築や構成変更にかかる時間を大幅に短縮し, 物理 GPU クラスタの手動構築に比べ約 1/11 の時間で作業が可能であるとわかった。(表 1)

6 関連研究

Chavan ら (2021) は PCI パススルーと vGPU の性能比較を行い, vGPU 環境で 7~9% のオーバーヘッドが発

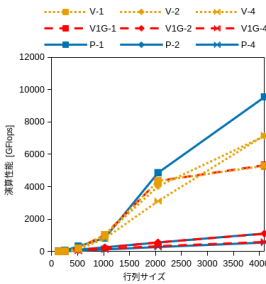


図 4 通信の割合が高いプログラムの性能変化

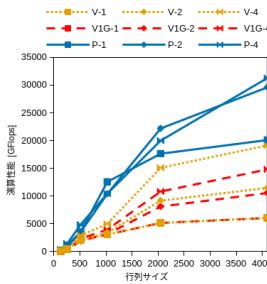


図 5 演算の割合が高いプログラムの性能変化

表 1 想定シナリオを完了する想定時間

項目	合計時間
Web UI システム	15 分 52 秒
物理クラスタ手動構築	3 時間 1 分 2 秒

生することを示した [2]。また, Hacker ら (2011) は HPC 教育における仮想クラスタの有用性を検証し, 柔軟な学習環境の提供が可能であることを報告した [3]。しかし, NVIDIA Ampere アーキテクチャを用いた vGPU 性能や仮想 GPU クラスタの運用に関する検証は不足している。本研究はその課題に対する評価を試みた。

7 おわりに

本研究では, NVIDIA A6000 における vGPU の性能評価に加え, 物理 GPU クラスタと仮想 GPU クラスタの特性を比較した。また, 仮想 GPU クラスタ構築システムの実装を行い, クラスタの構成変更にかかる時間を短縮した。

今後の課題としては, より大規模な GPU クラスタと仮想 GPU クラスタの性能比較, より高速な Ethernet, Infiniband, NVLink などのネットワークを利用した GPU クラスタとの比較, 他の方式の GPU 仮想化での評価などを検討している。

参考文献

[1] Nvidia virtual gpu software v16.0 through 16.9 - nvidia docs, [Online; accessed 2025-01-30]. [Online]. Available: <https://docs.nvidia.com/vgpu/16.0/index.html>.
[2] A. S. Chavan, A. S. Varal, V. S. Bhende, and M. Thalor, “Experimental analysis of dedicated gpu in virtual framework using vgpu,” in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2021, pp. 483–489. doi: 10.1109/ESCI50559.2021.9396984.
[3] T. Hacker, “Exploring the use of virtual machines and virtual clusters for high performance computing education.,” 2011. doi: 10.18260/1-2--17971.