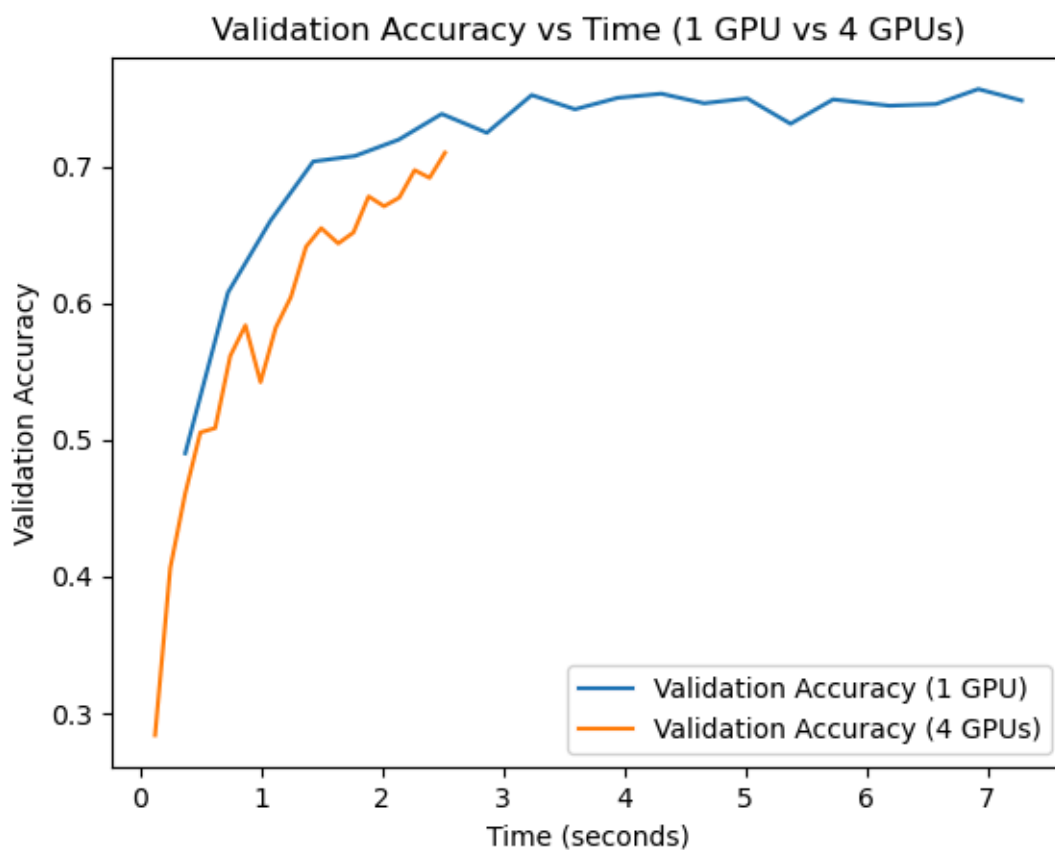


Distributed Deep Learning HW 2

Rajeev Atla

In this assignment, we trained 2 different versions of Resnet18 on the CIFAR10 dataset. A node of 4 Nvidia A100 GPUs were used on NSERC's Perlmutter supercomputer for the training. Please see the attached Python files for training details. (Pytorch Lightning's DDP strategy was used for ease of usage.)

The first version used data parallelism with 4 GPUs. We can plot the validation accuracy vs time.



The second version used model parallelism with 2 GPUs. The model was split up into 2 parts - 1 with 2.8 million parameters, and 1 with 8.4 million parameters.

