Machine Learning at TACC
Distributed Training

July 2024

PRESENTED BY:

Sikan Li

Research Engineering, Scalable Computational Intelligence

sli@tacc.utexas.edu

Distributed Training: Motivation

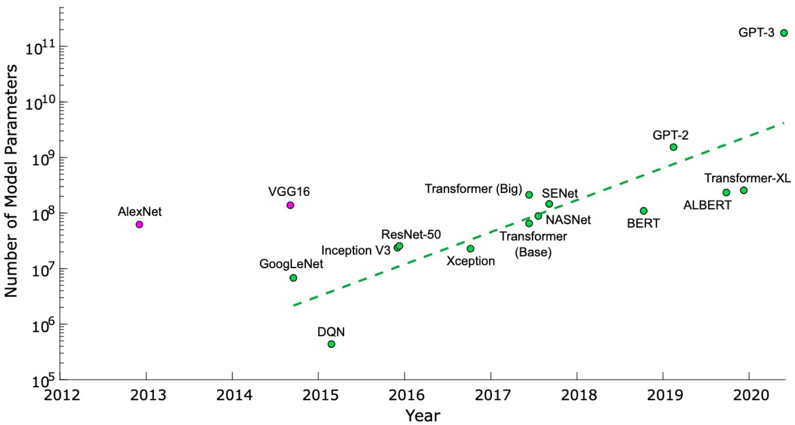


Figure 1. Number of parameters, i.e., weights, in recent landmark neural networks [1,2,31–43] (references dated by first release, e.g., on arXiv). The number of multiplications (not always reported) is not equivalent to the number of parameters, but larger models tend to require more compute power, notably in fully-connected layers. The two outlying nodes (pink) are AlexNet and VGG16, now considered over-parameterized. Subsequently, efforts have been made to reduce DNN sizes, but there remains an exponential growth in model sizes to solve increasingly complex problems with higher accuracy.

Image Credit

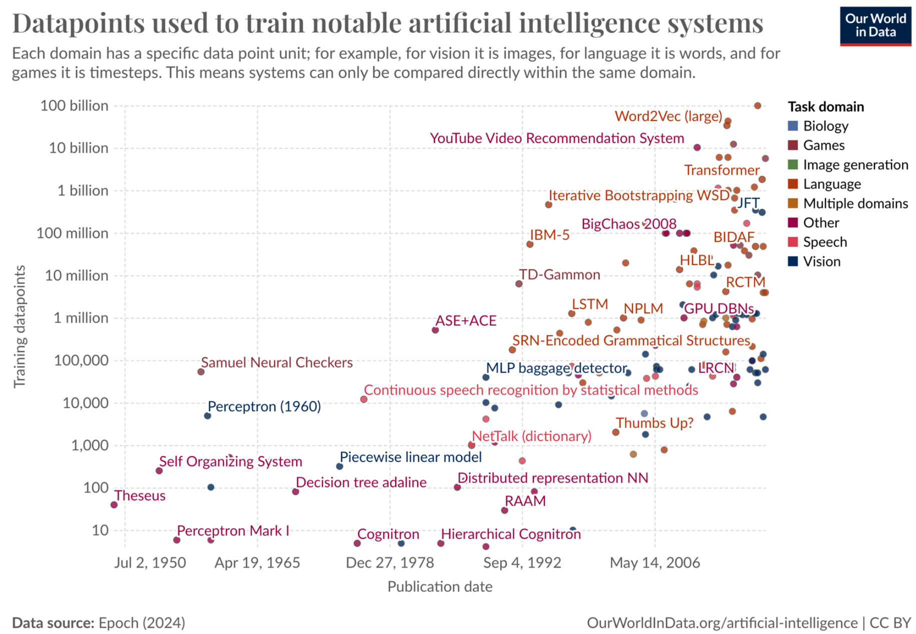


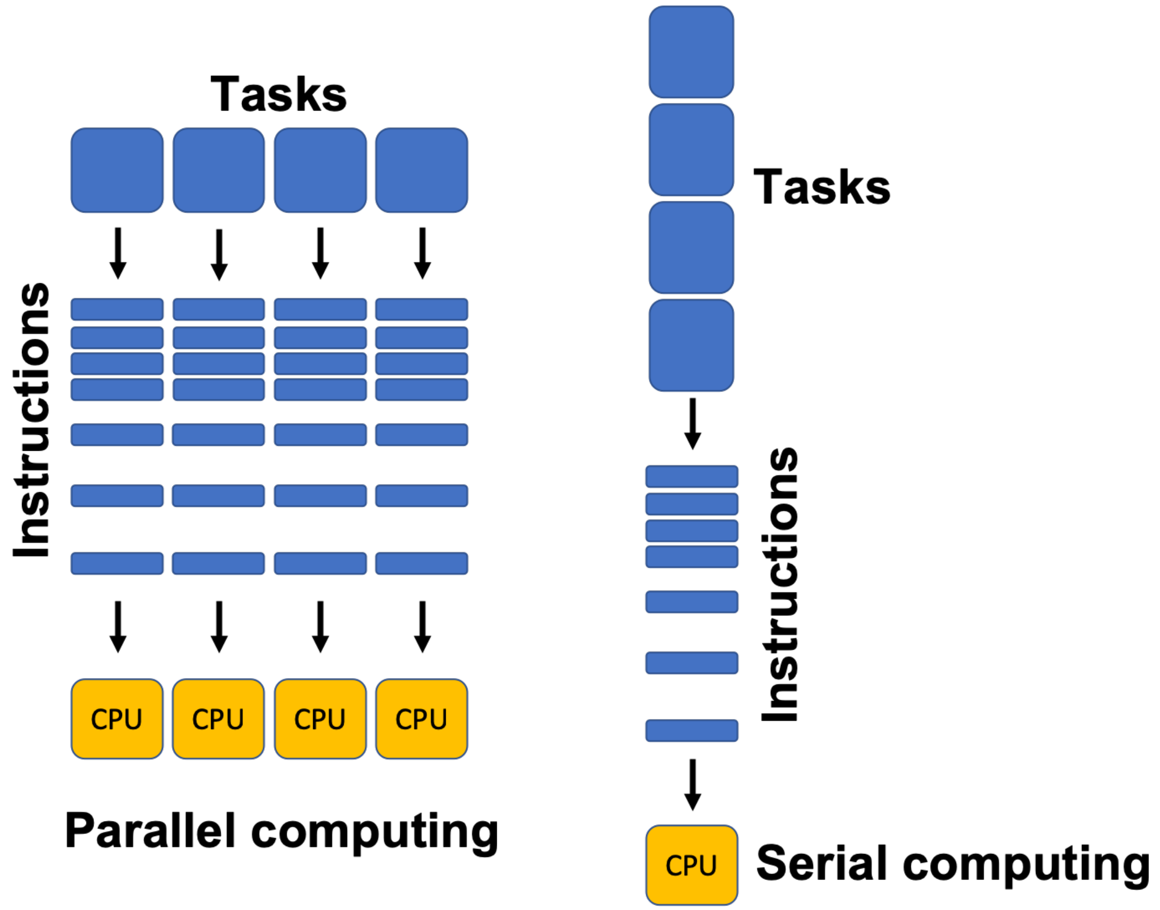
Image Credit

Outline

- Introduction to Parallel Computing on HPC
- Basics of Distributed Data Parallel
- Fault Tolerance & Torchrun
- MPI and SLURM
- Hands-on

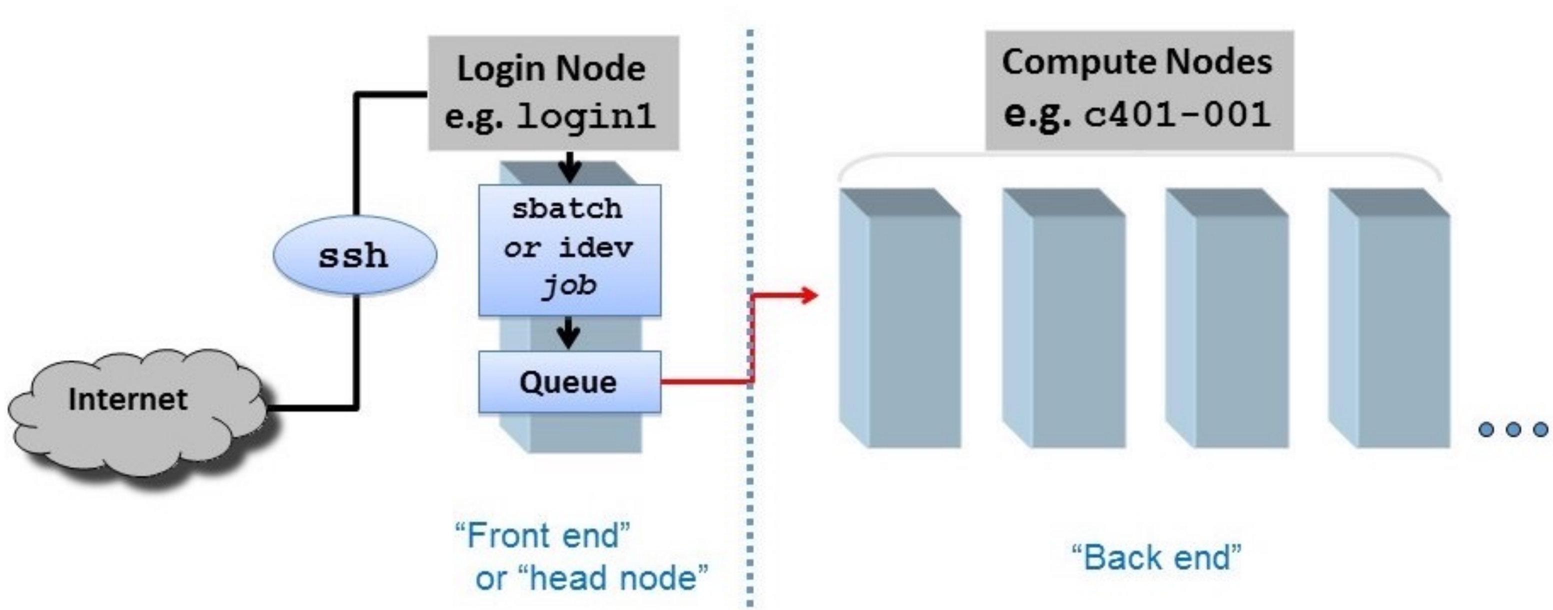
Introduction to parallel computing on High Performance Computing (HPC)

Introduction to parallel computing on HPC



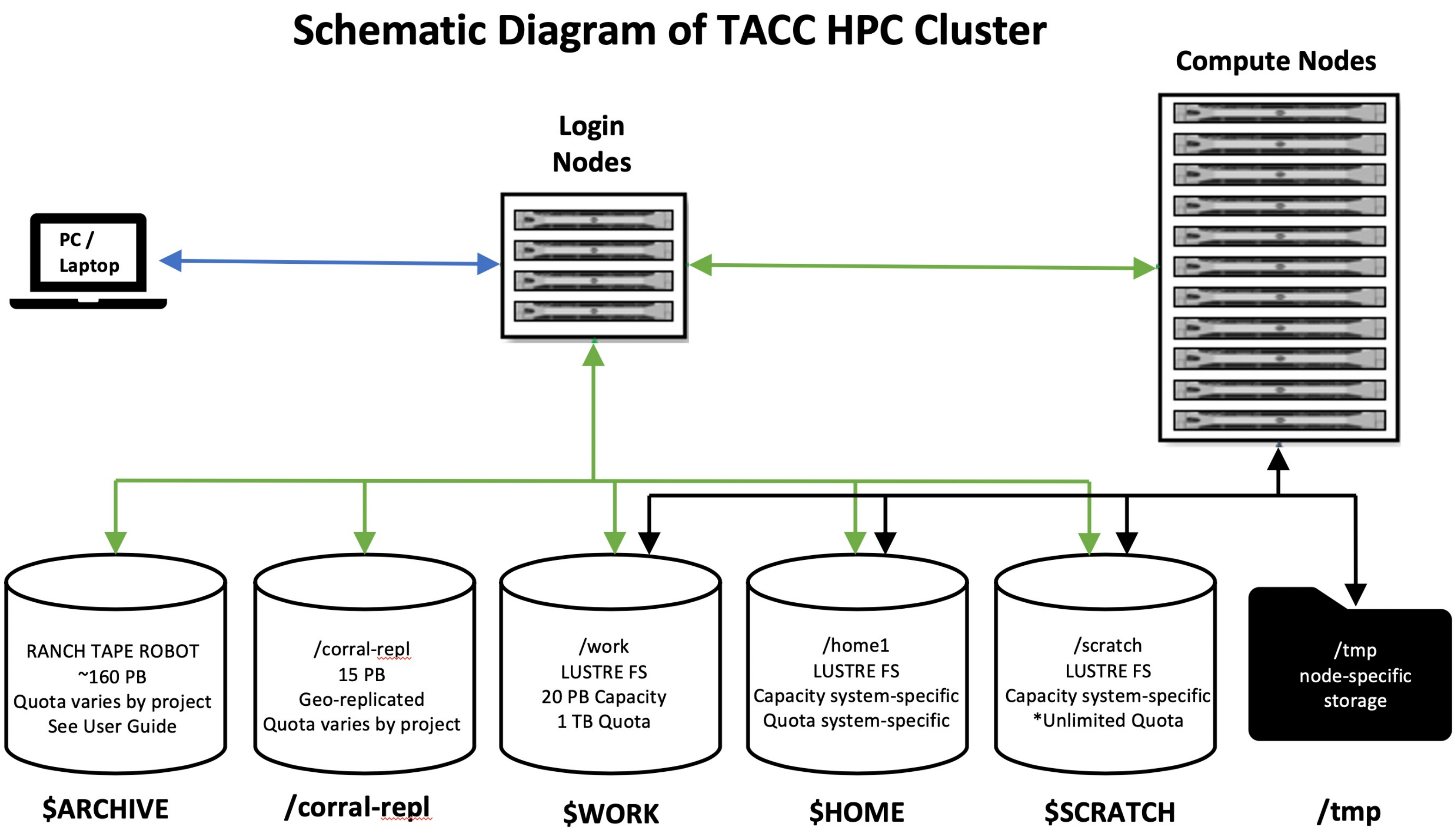
Serial vs. Parallel Computation

Introduction to Parallel Computing on HPC



Login vs Compute Nodes

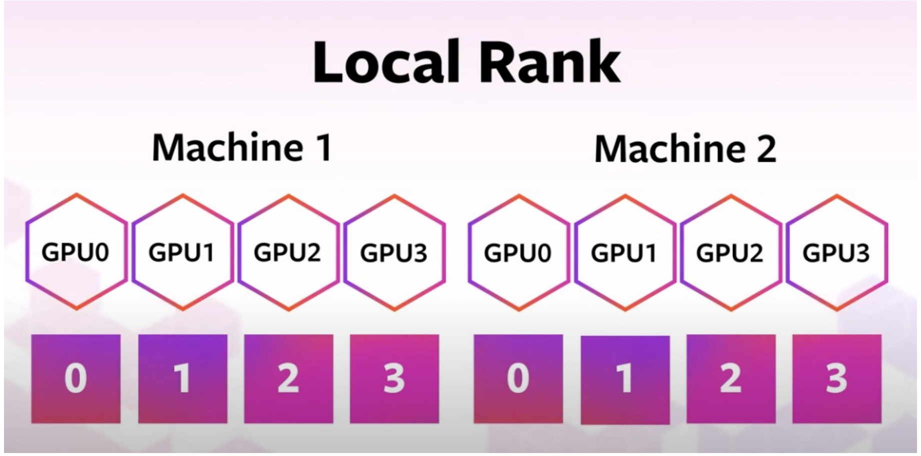
Introduction to Parallel Computing on HPC

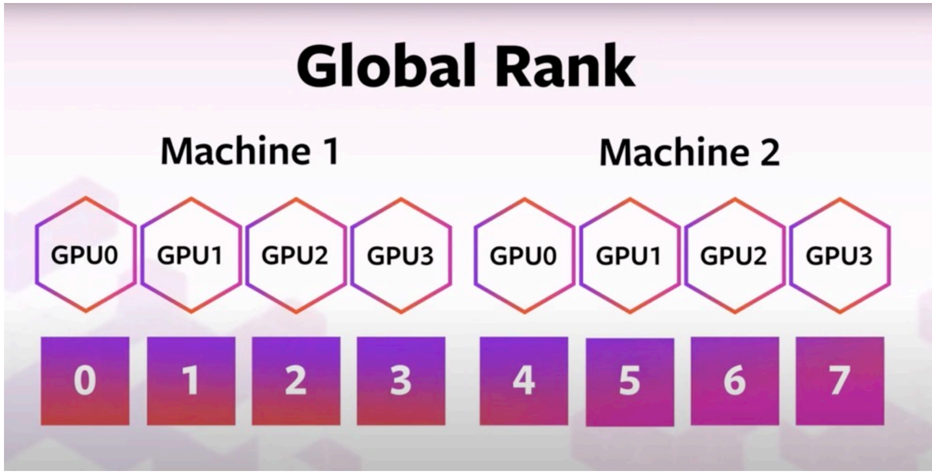


TACC

Introduction to Parallel Computing on HPC

- Local Rank & Global Rank
- World Size





Local Rank

Global Rank

Introduction to Parallel Computing on HPC

Challenge: Communication

A = B * C

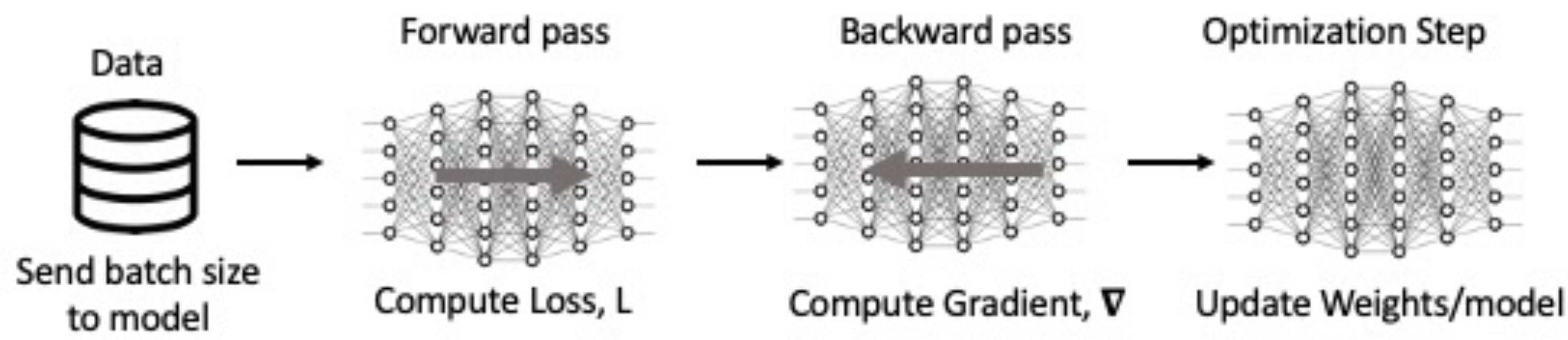
D = E * F

G = A * D

- Synchronization
- Data Movement
- Collective Computation

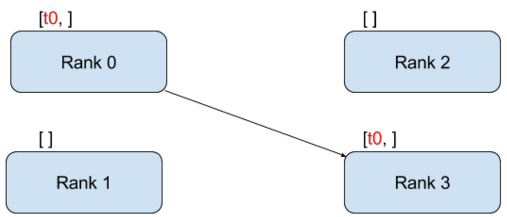
Basics of Distributed Data Parallel (DDP)

Basics of DDP



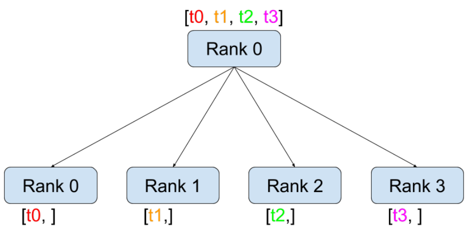
Visual representation of one step in training a neural network with one GPU.

Basics of DDP

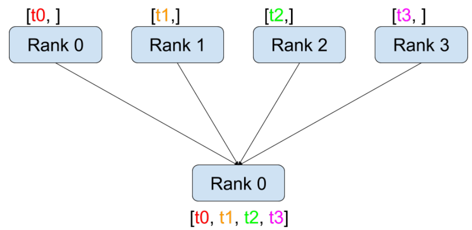


Point-to-Point Communication

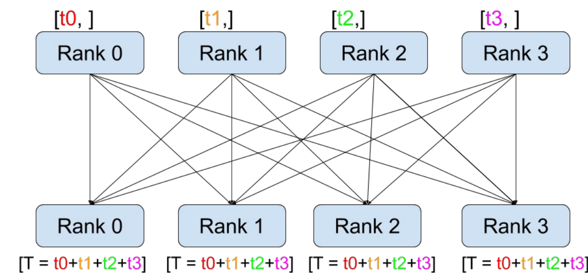
Basics of DDP



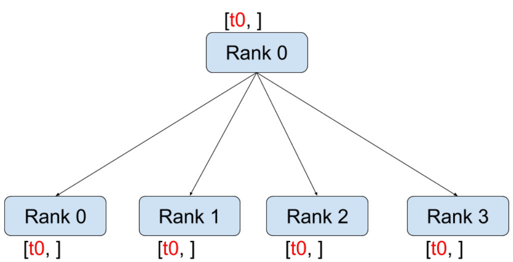
Scatter



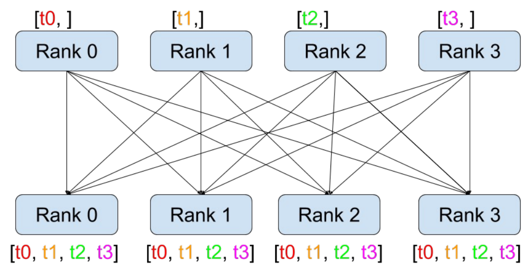
Gather



All-Reduce



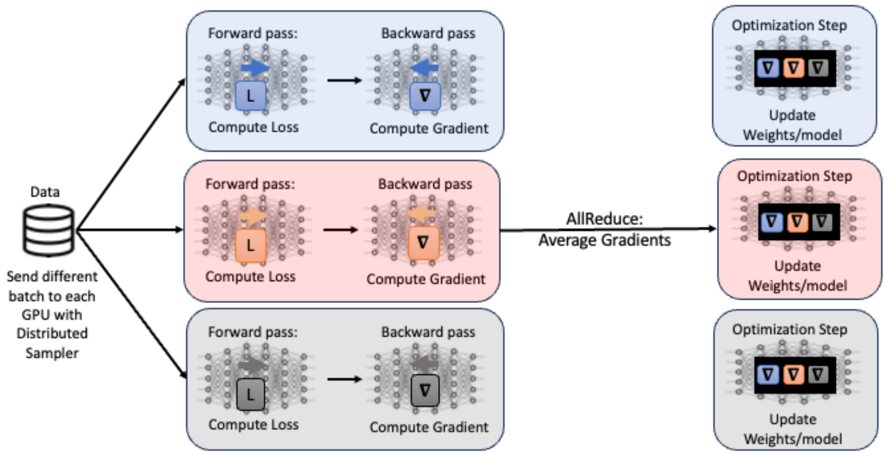
Broadcast



All-Gather

[Collective Communication]

Basics of DDP



Visual representation of one iteration of DDP using 3 GPUs.

- Synchronization
- Data Movement
- Collective Computation

Basics of DDP

PyTorch Distributed

DistributedDataParallel

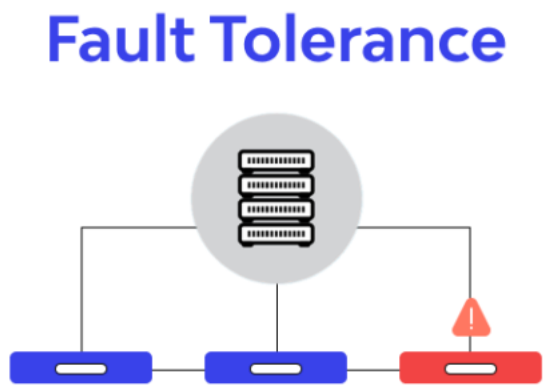
Basics of DDP

Communication backend for torch.distributed

- Gloo
- NVIDIA Collective Communication Library (NCCL)
- Message Passing Interface (MPI)

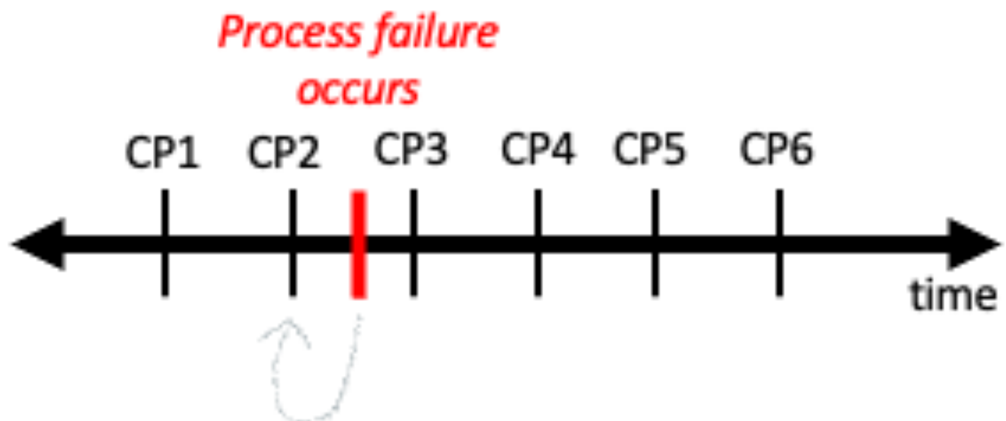
Fault Tolerance & Torchrun

Fault Tolerance & Torchrun



Fault Tolerance

Fault Tolerance & Torchrun



Visual of checkpointing. CP refers to a point in time when a checkpoint is saved.

Fault Tolerance & Torchrun

Torchrun:

torch.distributed.launch +

- worker failures
- environment variables
- elasticity

Message Passing Interface (MPI) & Simple Linux Utility for Resource Management (SLURM)

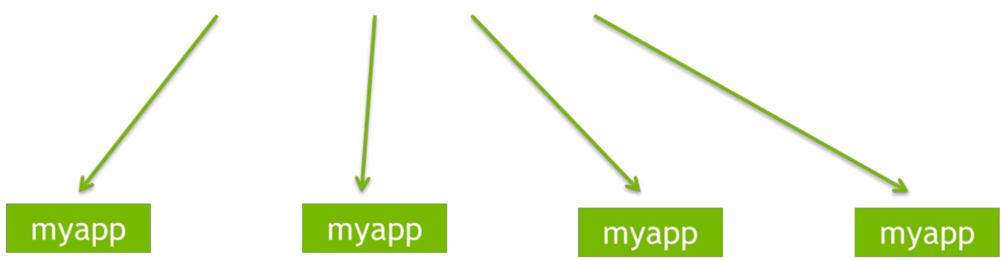
MPI & SLURM

Message Passing Interface (MPI)

MPI is a standardized and portable message-passing standard designed to function on parallel computing architectures.

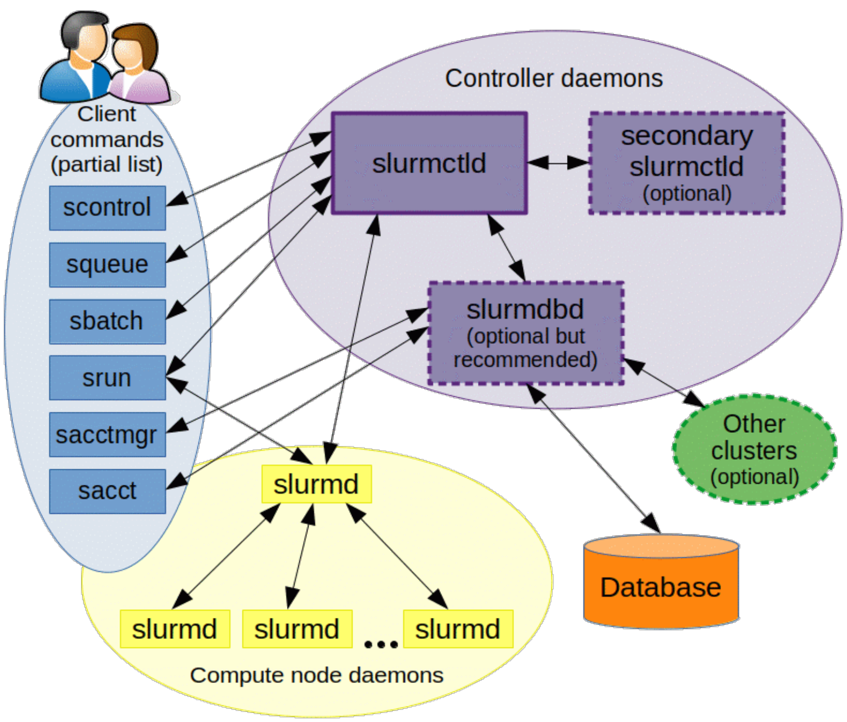
mpirun

```
mpirun -np 4 ./myapp <args>
```



MPI & SLURM

Simple Linux Utility for Resource Management (SLURM)



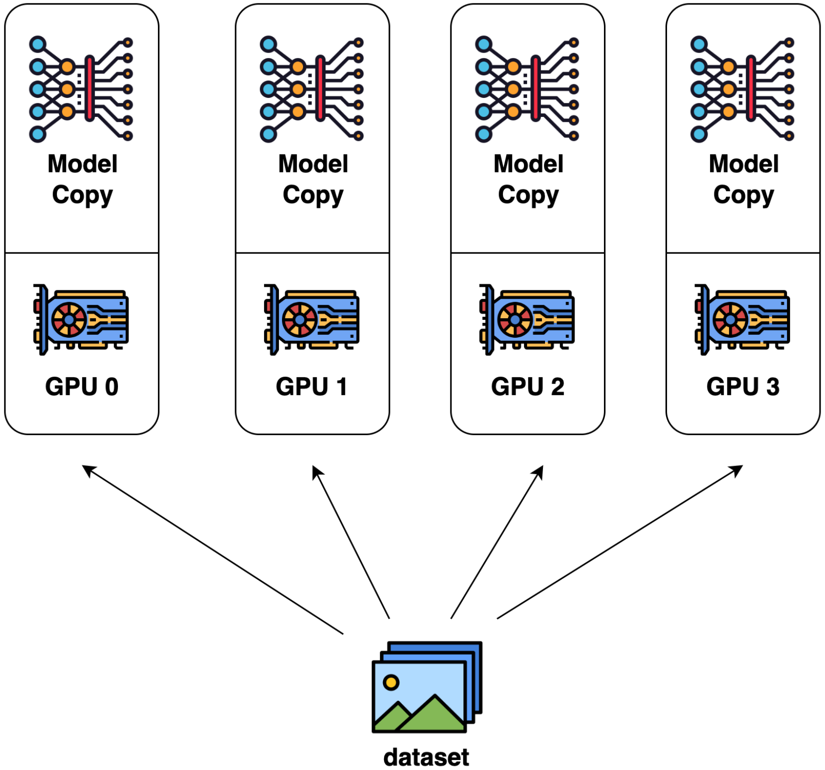
[Slurm components]

More Parallelization Techniques

- Data Parallel
- Model Parallel
- Pipeline Parallel

More Parallelization Techniques

Data Parallel



[Image Credit](#)

Parameters (Model) are duplicated on all nodes

More Parallelization Techniques

Pipeline Parallel

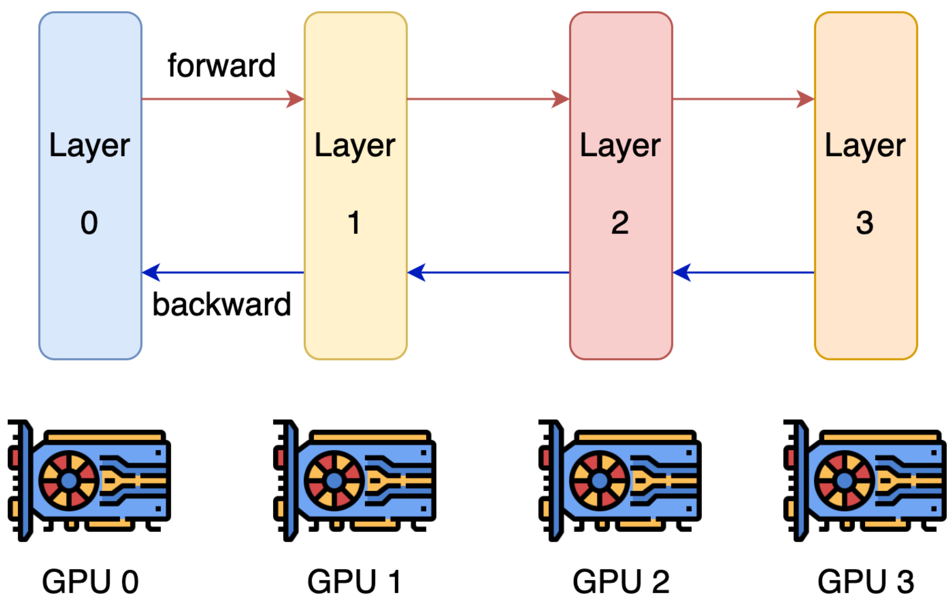
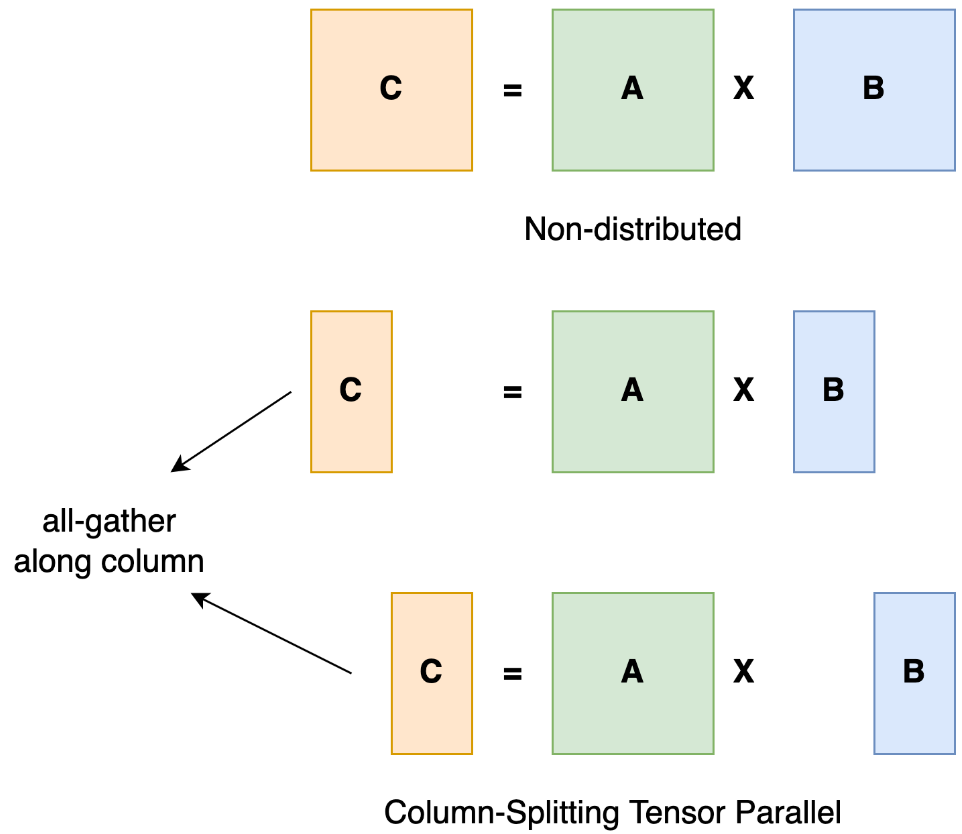


Image Credit

Weights are distributed to all nodes

More Parallelization Techniques

Model Parallel



Partitions the individual layers.

[Image Credit](#)

More Parallelization Techniques

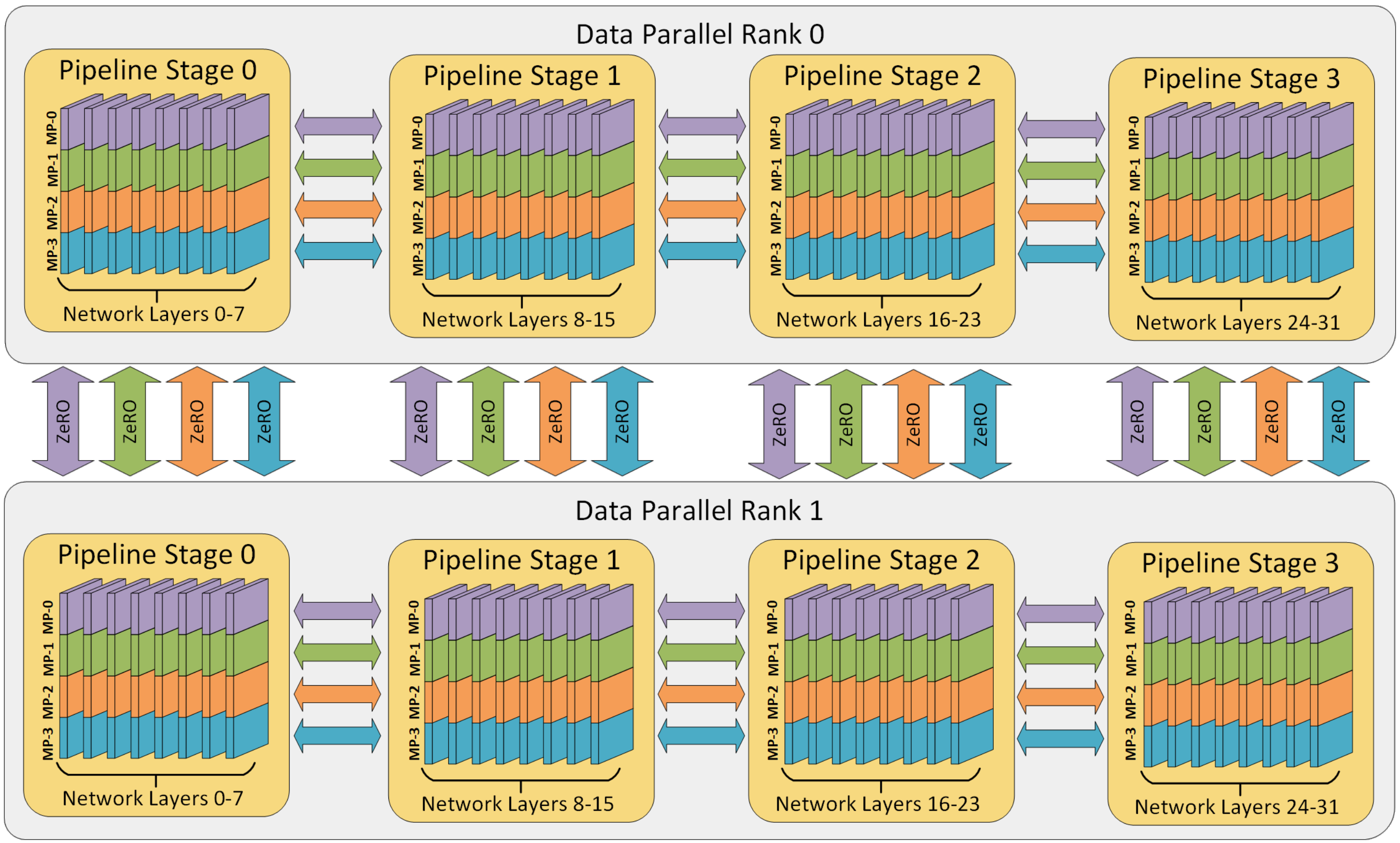


Image Credit

Hands-on Exercise

Hands-on Exercise

In the three notebooks, we will introduce how to parallelize the training process of a CNN classifier.

We will start with multiple GPUs on a single node, and then approach multi node distributed training.

Hands-on Exercise

Launch a Jupyter Notebook

- Resource: Frontera
- Project: Frontera-Training
- Session Type: Jupyter Notebook
- Reservation ID: ML_Institute_Thu
- Queue: rtx
- Time Limit: 04:00:00