Data Parallel DNN Training with Tensorflow

High Performance Machine Learning
CS 5463

Buddhi Ashan M. K.

Spring 2025

Overview

- TensorFlow API to distribute training across multiple GPUs, multiple machines, or TPUs
  - `tf.distribute.Strategy`
- Supports
  - high-level API methods via Keras
  - Custom training loops

Class Code Repository

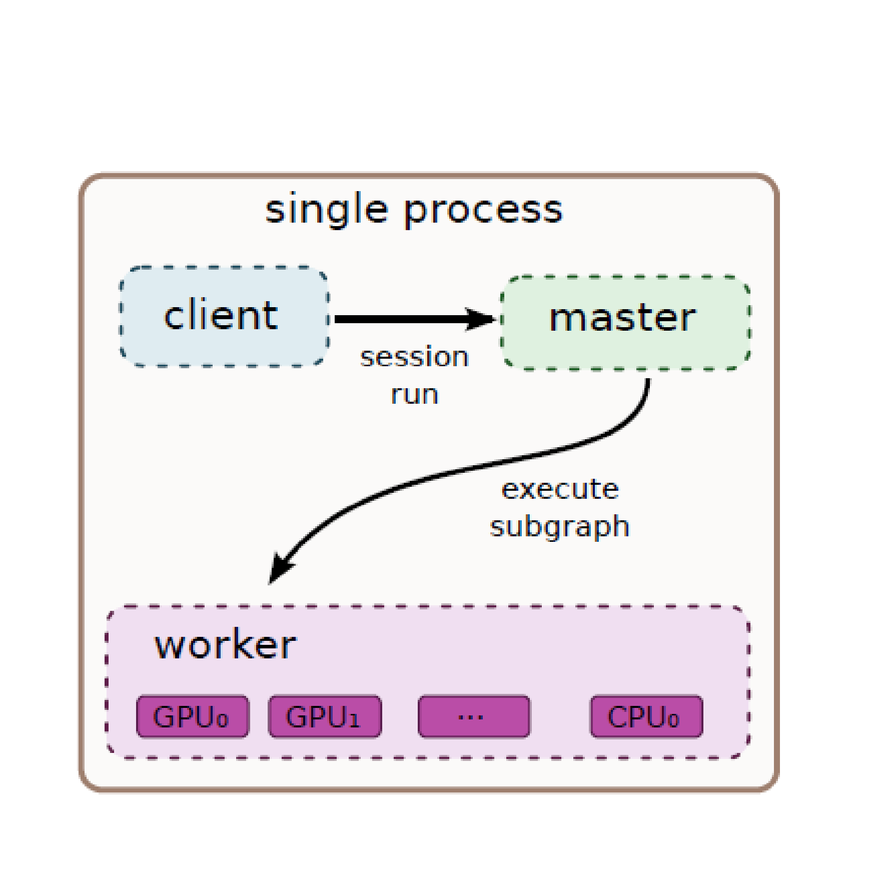
- All code files we are using in the class can be found at the following git repo:
  - https://github.com/buddhi1/pdc-class.git
- Clone the repo
  - `git clone https://github.com/buddhi1/pdc-class.git`
- Or download as a zip file

Types of Strategies

- Synchronous
  - `MirroredStrategy`
  - `MultiWorkerMirroredStrategy`
  - `CentralStorageStrategy`
- Asynchronous training
  - `ParameterServerStrategy`

- Support different platforms including multiple GPUs, multiple machines, and TPUs

Single node Multi GPU DNN Training



Abadi, Martın, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. n.d. "TensorFlow: A System for Large-Scale Machine Learning."

MirroredStrategy

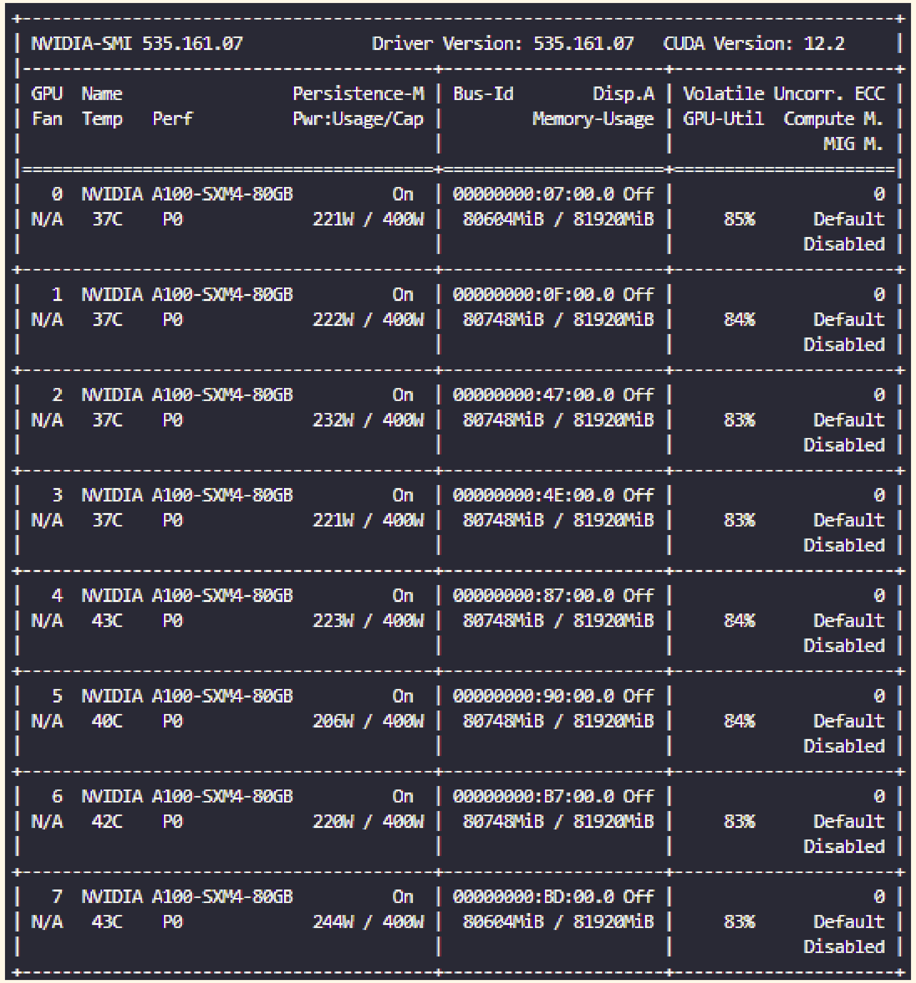
- Synchronous distributed training on multiple GPUs on one machine.
- The model replicated across all devices.
- Together, these variables form a single conceptual variable called ***MirroredVariable***.
- These variables are kept in sync with each other by applying identical updates.

Arc

```
srun -p gpu2v100 -gres=gpu:2 -N 1 -n 64 -t 0:30:0 --pty bash
ml  anaconda3
conda create --name tf-gpu tensorflow-gpu
conda activate tf-gpu
```

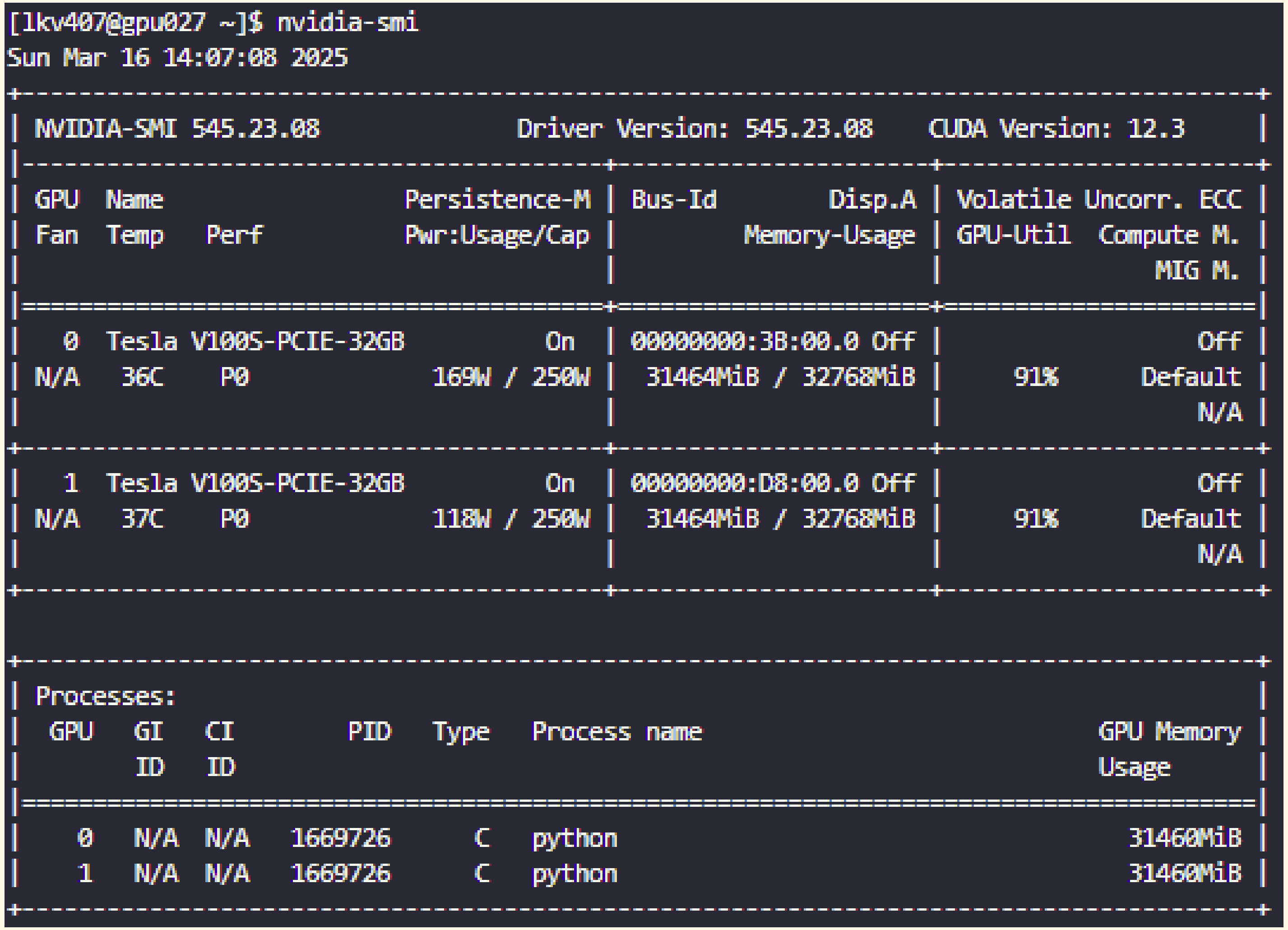
- All commands can be found in the commandsForArc.txt

dgx with Eight GPUs



```
+---------------------------------------------------------------------------------------+
| NVIDIA-SMI 535.161.07             Driver Version: 535.161.07   CUDA Version: 12.2     |
|-----------------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|=========================================+======================+======================|
|   0  NVIDIA A100-SXM4-80GB          On  | 00000000:07:00.0 Off |                    0 |
| N/A   37C    P0             221W / 400W |  80604MiB / 81920MiB |     85%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   1  NVIDIA A100-SXM4-80GB          On  | 00000000:0F:00.0 Off |                    0 |
| N/A   37C    P0             222W / 400W |  80748MiB / 81920MiB |     84%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   2  NVIDIA A100-SXM4-80GB          On  | 00000000:47:00.0 Off |                    0 |
| N/A   37C    P0             232W / 400W |  80748MiB / 81920MiB |     83%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   3  NVIDIA A100-SXM4-80GB          On  | 00000000:4E:00.0 Off |                    0 |
| N/A   37C    P0             221W / 400W |  80748MiB / 81920MiB |     83%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   4  NVIDIA A100-SXM4-80GB          On  | 00000000:87:00.0 Off |                    0 |
| N/A   43C    P0             223W / 400W |  80748MiB / 81920MiB |     84%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   5  NVIDIA A100-SXM4-80GB          On  | 00000000:90:00.0 Off |                    0 |
| N/A   40C    P0             206W / 400W |  80748MiB / 81920MiB |     84%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   6  NVIDIA A100-SXM4-80GB          On  | 00000000:B7:00.0 Off |                    0 |
| N/A   42C    P0             220W / 400W |  80748MiB / 81920MiB |     83%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
|   7  NVIDIA A100-SXM4-80GB          On  | 00000000:BD:00.0 Off |                    0 |
| N/A   43C    P0             244W / 400W |  80604MiB / 81920MiB |     83%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+
```

Arc: 1 GPU



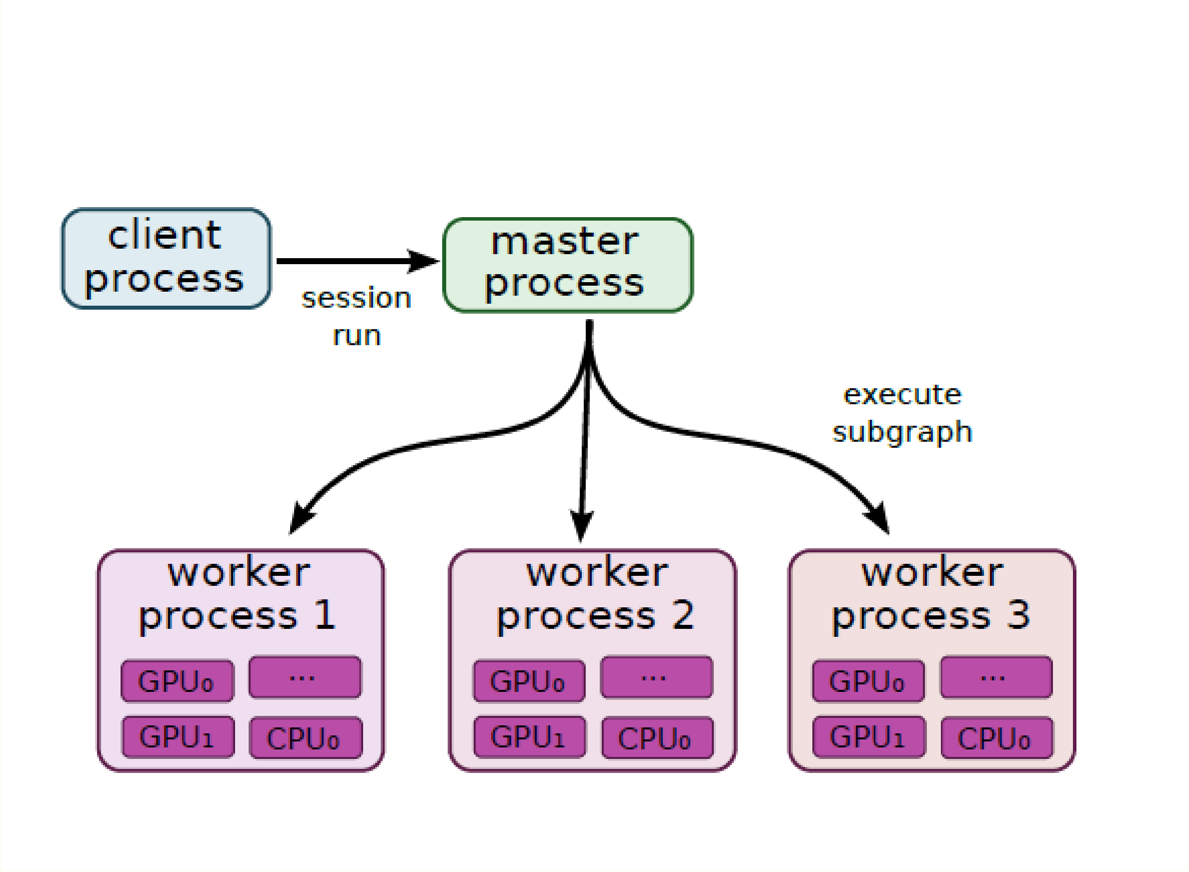
```
[lkv407@gpu027 ~]$ nvidia-smi
Sun Mar 16 14:07:08 2025
+---------------------------------------------------------------------------------------+
| NVIDIA-SMI 545.23.08              Driver Version: 545.23.08    CUDA Version: 12.3     |
|-----------------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|=========================================+======================+======================|
|   0  Tesla V100S-PCIE-32GB          On  | 00000000:3B:00.0 Off |                  Off |
| N/A   36C    P0             169W / 250W |  31464MiB / 32768MiB |     91%      Default |
|                                         |                      |                  N/A |
+-----------------------------------------+----------------------+----------------------+
|   1  Tesla V100S-PCIE-32GB          On  | 00000000:D8:00.0 Off |                  Off |
| N/A   37C    P0             118W / 250W |  31464MiB / 32768MiB |     91%      Default |
|                                         |                      |                  N/A |
+-----------------------------------------+----------------------+----------------------+

+---------------------------------------------------------------------------------------+
| Processes:                                                                            |
|  GPU   GI   CI        PID   Type   Process name                            GPU Memory |
|        ID   ID                                                             Usage      |
|=======================================================================================|
|    0   N/A  N/A   1669726      C   python                                    31460MiB |
|    1   N/A  N/A   1669726      C   python                                    31460MiB |
+---------------------------------------------------------------------------------------+
```

Cross Device Communication

- `tf.distribute.HierarchicalCopyAllReduce`
- `tf.distribute.ReductionToOneDevice`
- `tf.distribute.NcclAllReduce` (default)

```
mirrored_strategy = tf.distribute.MirroredStrategy(
        cross_device_ops=tf.distribute.HierarchicalCopyAllReduce())
```

Distributed DNN Training



Abadi, Martın, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. n.d. "TensorFlow: A System for Large-Scale Machine Learning."

MultiWorkerMirroredStrategy

- Synchronous distributed training across multiple workers, each with potentially multiple GPUs.
- Creates copies of all variables in the model on each device across all workers

Cross Device Communication

- **`communicationImplementation.RING`**: RPC-based and supports both CPUs and GPUs.
- **`communicationImplementation.NCCL`**: NCCL and provides state-of-art performance on GPUs but it doesn't support CPUs.
- **`collectiveCommunication.AUTO`**: defers the choice to Tensorflow

```
communication_options = tf.distribute.experimental.CommunicationOptions(

         implementation=tf.distribute.experimental.CommunicationImplementation.N
CCL)
strategy = tf.distribute.MultiWorkerMirroredStrategy(
                  communication_options=communication_options)
```

Google Colab

- https://colab.research.google.com/drive/1VBfV3D9Sa4G3o_Uys4egCFLIO-jNlQV4?usp=sharing

Arc

- `srun -p compute1 -N 2 -n 128 -t 0:30:0 --pty bash`
- `ml  anaconda3`
- `conda create --name tf-gpu tensorflow-gpu`
- `conda activate tf-gpu`
- How to get the list of available nodes:
  - `echo $SLURM_NODELIST`
- `ssh <other node id>`
- `ml  anaconda3`
- `conda create --name tf-gpu tensorflow-gpu`
- `conda activate tf-gpu`