**Guidance to run PyTorch BERT-Large PreTraining on Intel Max 1100 GPUs**

Login to ACES cluster and run the commands below.

$cd $SCRATCH

$mkdir pvc-benchmarks

$cd pvc-benchmarks

$git clone -b r2.12.1 <https://github.com/IntelAI/models.git>

$ml purge

$ml GCCcore/11.2.0 Python/3.9.6

$python3 -m venv bert-large-pt-training-trial

$source bert-large-pt-training-trial/bin/activate

$pip install torch==2.1.0.post0 torchvision==0.16.0.post0 torchaudio==2.1.0.post0

intel\_extension\_for\_pytorch==2.1.20+xpu oneccl-bind-pt==2.1.200 deepspeed==0.14.0

--extra-index-url <https://pytorch-extension.intel.com/release-whl-aitools/>

$cd models

$chmod 755 quickstart/language\_modeling/pytorch/bert\_large/training/gpu/\*.sh

$pip install -r

models/language\_modeling/pytorch/bert\_large/training/gpu/requirements.txt

$wget https://s3.amazonaws.com/models.huggingface.co/bert/bert-base-uncased-

vocab.txt -O models/language\_modeling/pytorch/bert\_large/training/gpu/data/vocab.txt

$deactivate

# create a slurm job file test\_pytorch\_bert\_large.slurm and copy and paste the content below to it.

$ vim test\_pytorch\_bert\_large.slurm

#!/bin/bash

##NECESSARY JOB SPECIFICATIONS

#SBATCH --job-name=<your\_job\_name>

#SBATCH --time=2:00:00 # the wallclock time for a job

#SBATCH --nodes=1 # total number of nodes

#SBATCH --ntasks=1 # total number of processes

#SBATCH --mem=60G

#SBATCH --output=<your\_job\_output>.%j # output of your slurm job

#SBATCH --gres=gpu:pvc:1 # for 2 gpus, set --gres=gpu:pvc:2

#SBATCH --partition=pvc # partition should be pvc for intel gpus

echo "Hostname: $(hostname)"

# Print the node name

echo "Node name: $SLURMD\_NODENAME"

sinfo -N -p pvc -o "%8n %10f %G"

# This command is used to get stats of Max GPU utilization

xpumcli dump -m 0,1,2,3,4,5,6,17,18,19,20,35 > <your\_job>\_Max\_GPU\_stats.log &

# This command is used to get stats of how many CPU cores are being used

watch -n 5 ps -u $USER > <your\_job>\_Max\_CPU\_stats.log &

# load all the necessary modules module purge

ml purge

ml GCCcore/11.2.0 Python/3.9.6

source $SCRATCH/pvc-benchmarks/bert-large-pt-training-trial/bin/activate

source /sw/hprc/sw/oneAPI/2024.0/setvars.sh

# set environment variales

export DATASET\_DIR=/scratch/data/pytorch-language-modelling-datasets/mlcommons-dataset

export PROCESSED\_DATASET\_DIR=/scratch/data/pytorch-language-modelling-datasets/processed\_mlcommons\_dataset

export BATCH\_SIZE=32

export NUM\_ITERATIONS=10000

export TF\_ENABLE\_ONEDNN\_OPTS=0

cd $SCRATCH/pvc-benchmarks/models

# a tool on ACES cluster to get graph from stats of GPU and CPU utilization

jobstats &

export OUTPUT\_DIR=$SCRATCH/pvc-benchmarks/output\_logs

bash quickstart/language\_modeling/pytorch/bert\_large/training/gpu/bf16\_training\_plain\_format.sh

jobstats

$sbatch test\_pytorch\_bert\_large.slurm