[TO BE VALIDATED BY ENGINEERING]

Here are the instructions to accelerate TensorFlow models on AWS with the Intel® OpenVINOTM add-on for TensorFlow

1. Launch the Deep Learning AMI EC2 instance Ubuntu 18.04 Version 41.0  
     
    Graphical user interface, text, application, email

   Description automatically generated
2. Choose one of the C5 instances –these are optimized for inference. The larger the instance , the faster the inference.

A screenshot of a computer

Description automatically generated with low confidence

1. Then download the \*.pem files for keys on your laptop.   
   chmod 400 the \*.pem key.
2. Get the public IP address of your instance

Graphical user interface, text, application

Description automatically generated

1. Wait for the instance to finish initializing and be fully running and functional
2. ssh -i \*.pem <IP-addr-of-your-instance>
   1. scp -i \*.pem <source-file> <IP-addr-of-your-instance>:/tmp

SSH and SCP with the AWS instance should be working (Note: It worked seamlessly on our team Intel provided AWS account, but developers might need to configure networking to enable this.)

1. git clone https://github.com/openvinotoolkit/openvino\_tensorflow.git  
     
   OR just download zip of the whole repo to your local laptop and then scp it to the AWS instance (you might need to scp to :/tmp for permission reasons, and then ssh to the instance then copy the zip to your home directory)
2. **ubuntu@ip-10-0-0-123**:**~**$python3 -m venv myenv
3. **ubuntu@ip-10-0-0-123**:**~**$ source myenv/bin/activate
4. (myenv) **ubuntu@ip-10-0-0-123**:**~**$ pip install --upgrade pip
5. (myenv) **ubuntu@ip-10-0-0-123**:**~**$ pip install tensorflow==2.2.2
6. (myenv) **ubuntu@ip-10-0-0-123**:**~**$ pip install -U --index-url https://test.pypi.org/simple/ --extra-index-url https://pypi.org/simple/ openvino-tensorflow-addon-abi0
7. (myenv) **ubuntu@ip-10-0-0-123**:**~**$ python3 -c "import tensorflow as tf; print('TensorFlow version: ',tf.\_\_version\_\_); import openvino\_tensorflow; print(openvino\_tensorflow.\_\_version\_\_)"

TensorFlow version: 2.2.2

openvino tensorflow add-on version: 0.5.0

nGraph version used for this build: b'0.0.0+a8a6e8f'

TensorFlow version used for this build: v2.2.2-1-g876c0a59768

CXX11\_ABI flag used for this build: 0

openvino tensorflow add-on built with Grappler: False

1. (myenv) **ubuntu@ip-10-0-0-123**:**~**$ cd openvino\_tensorflow-master/
2. (myenv) **ubuntu@ip-10-0-0-123**:**~/openvino\_tensorflow-master**$ curl -L "https://storage.googleapis.com/download.tensorflow.org/models/inception\_v3\_2016\_08\_28\_frozen.pb.tar.gz" | tar -C ./examples/data -xz

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

100 84.5M 100 84.5M 0 0 29.0M 0 0:00:02 0:00:02 --:--:-- 29.0M

1. (myenv) **ubuntu@ip-10-0-0-123**:**~/openvino\_tensorflow-master**$ python3 examples/classification\_sample.py

2021-03-16 23:38:39.107565: I tensorflow/stream\_executor/platform/default/dso\_loader.cc:44] Successfully opened dynamic library libcuda.so.1

2021-03-16 23:38:39.609060: E tensorflow/stream\_executor/cuda/cuda\_driver.cc:313] failed call to cuInit: CUDA\_ERROR\_NO\_DEVICE: no CUDA-capable device is detected

2021-03-16 23:38:39.609158: I tensorflow/stream\_executor/cuda/cuda\_diagnostics.cc:156] kernel driver does not appear to be running on this host (ip-10-0-0-123): /proc/driver/nvidia/version does not exist

2021-03-16 23:38:39.610135: I tensorflow/core/platform/cpu\_feature\_guard.cc:143] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 AVX512F FMA

2021-03-16 23:38:39.624320: I tensorflow/core/platform/profile\_utils/cpu\_utils.cc:102] CPU Frequency: 2999995000 Hz

2021-03-16 23:38:39.627801: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0x7f827c000b20 initialized for platform Host (this does not guarantee that XLA will be used). Devices:

2021-03-16 23:38:39.627822: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version

Available Backends:

CPU

Inference time in ms: 7.504463

military uniform 0.8343049

mortarboard 0.021869553

academic gown 0.010358133

pickelhaube 0.008008199

bulletproof vest 0.0053509558