## TABLE OF CONTENTS

Chapter 1. Deep Learning Profiler 20.03 Release Notes ................................................ 1  
Chapter 2. Deep Learning Profiler 20.02 Release Notes ................................................ 3  
Chapter 3. Deep Learning Profiler 20.01 Release Notes ................................................ 5  
Chapter 10. Deep Learning Profiler 19.06 Release Notes ........................................... 15
Chapter 1.
DEEP LEARNING PROFILER 20.03 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Driver Requirements
Release 20.03 is based on NVIDIA CUDA 10.2.89, which requires NVIDIA Driver release 440.33.01. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 396, 384.111+, 410, 418.xx, or 440.30. The CUDA driver’s compatibility package only supports particular drivers. For a complete list of supported drivers, see the CUDA Application Compatibility topic. For more information, see CUDA Compatibility and Upgrades.

New Features
The key features of DLProf v0.10.0 / r20.03 are:

▶ Released in the TensorFlow 20.03 NGC container.
▶ Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, and Nsight Systems 2020.1.1.
▶ Expert Systems feature that analyzes performance results, looks for common performance issues, and suggests recommended fixes that may improve performance.
▶ Support for additional domains from custom NVTX markers.
   ▶ Reports are generated for the domain specified using markers.
   ▶ Data is aggregated only from NVTX markers in the same domain.
Passing a Graphdef is now optional. User can specify a Graphdef with --graphdef or set it to auto for a TensorBoard graph event file to be created.

System information is gathered in the background and is exposed in the summary report, database, and TensorBoard event files.

Consistent command line arguments.

Known Issues

This software is only accessible in the NGC TensorFlow container.

This software is only supported for TensorFlow 1.15 and TensorBoard 1.15.

The following command line options have been changed.

- `--in_nsys_db_filename` is now `--nsys_database`
- `--in_saved_model` was removed
- `--nsys_base_output_name` is now `--nsys_base_name`

Resolved Issues

- Fixed issue with XLA kernels and nodes not being aggregated correctly.
Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Driver Requirements
Release 20.02 is based on NVIDIA CUDA 10.2.89, which requires NVIDIA Driver release 440.30.01. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 396, 384.111+, 410, 418.xx, or 440.30. The CUDA driver's compatibility package only supports particular drivers. For a complete list of supported drivers, see the CUDA Application Compatibility topic. For more information, see CUDA Compatibility and Upgrades.

New Features
The key features of DLProf v0.9.0 / r20.02 are:

- Released in the TensorFlow 20.02 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, and Nsight Systems 2020.1.1.
- Added --delay and --duration options that will delay when the profiler will start and terminate the profiler after a set duration.
- Partial support for custom NVTX ranges and domains.
  - Can profile tensorflow models that use the NVTX Plugin.
  - Can select which domain(s) to use in generated report(s).
- New group node report that shows the aggregated times for each group node.
- Removed original TensorBoard GPU Summary Plugin.
Added pie charts and line series to the new TensorBoard DLProf Plugin.

**Known Issues**

- XLA cluster mapping in the TensorBoard Graph plugin is not supported in the 20.02 Tensorflow container.
- This software is only accessible in the NGC TensorFlow container.
- This software is only supported by TensorFlow 1.15.

**Resolved Issues**

- Iteration reports are now properly sorted.
- `--force` will work correctly with auto generated graphdef files.
Chapter 3. 
DEEP LEARNING PROFILER 20.01 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Driver Requirements
Release 20.01 is based on NVIDIA CUDA 10.2.89, which requires NVIDIA Driver release 440.30.01. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 396, 384.111+, 410, 418.xx, or 440.30. The CUDA driver’s compatibility package only supports particular drivers. For a complete list of supported drivers, see the CUDA Application Compatibility topic. For more information, see CUDA Compatibility and Upgrades.

New Features
The key features of DLProf v0.8.0 / r20.01 are:

- Released in the TensorFlow 20.01 NGC container
- Latest DLProf build is based on TensorFlow 1.15.0, TensorBoard 1.15.0, and Nsight Systems 2019.6.1
- Support for Tensorflow 1.15 and TensorBoard 1.15
- New DLProf Plugin for TensorBoard.
  - Currently, there is a BETA release of the plugin, and it is previewed along with the original GPU Summary Panel
  - Updated Summary page with new key metrics, including TC Utilization and GPU Idle %
  - Inclusion of Expert Systems feedback in a panel on the Summary page
Known Issues

- XLA cluster mapping in the TensorBoard Graph plugin is not supported in 20.01 Tensorflow container
- This software is only accessible in the NGC TensorFlow container
- This software is only supported by TensorFlow 1.15
Chapter 4.
DEEP LEARNING PROFILER 19.12 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Driver Requirements
Release 19.12 is based on NVIDIA CUDA 10.2.89, which requires NVIDIA Driver release 440.30. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 396, 384.111+, 410 or 418.xx. The CUDA driver's compatibility package only supports particular drivers. For a complete list of supported drivers, see the CUDA Application Compatibility topic. For more information, see CUDA Compatibility and Upgrades.

New Features
The key features of DLProf v0.7.0 / r19.12:

- Released in the TensorFlow 19.12 NGC container
- Latest DLProf build is based on TensorFlow 1.15.0, TensorBoard 1.15.0, and Nsight Systems 2019.6.1
- Support for Tensorflow 1.15 and TensorBoard 1.15
- Initial Expert Systems utility. DLProf now has an alpha version of Expert Systems that will analyze the profile results and provide recommendations on how to improve the training performance and profiling experience.
Known Issues

- XLA cluster mapping in the TensorBoard Graph plugin is not supported in 19.12 Tensorflow container
- This software is only accessible in the NGC TensorFlow container
- This software is only supported by TensorFlow 1.15
Chapter 5.
DEEP LEARNING PROFILER 19.11 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Driver Requirements
Release 19.11 is based on NVIDIA CUDA 10.2.89, which requires NVIDIA Driver release 440.30. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 396, 384.111+, 410 or 418.xx. The CUDA driver's compatibility package only supports particular drivers. For a complete list of supported drivers, see the CUDA Application Compatibility topic. For more information, see CUDA Compatibility and Upgrades.

New Features
The key features of DLProf v0.6.0 / r19.11:

- Released in the TensorFlow 19.11 NGC container
- Latest DLProf build is based on TensorFlow 1.15.0, TensorBoard 1.15.0, and Nsight Systems 2019.5.2
- Support for Tensorflow 1.15 and TensorBoard 1.15
- Updated time aggregation, improving reported time metrics

Known Issues

- XLA cluster mapping in the TensorBoard Graph plugin is not supported in 19.11 Tensorflow container
- This software is only accessible in the NGC TensorFlow container.
This software is only supported by TensorFlow 1.15
Chapter 6.
DEEP LEARNING PROFILER 19.10 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

New Features
The key features of DLProf v0.5.0/r19.10:

- GraphDef file generation
  - The GraphDef file is now automatically generated whenever you invoke DLProf.
  - You can still specify a pre-generated GraphDef file.
- Full XLA Support
  - You can now profile XLA compiled models to get maximum performance and see the profile output.
  - Reports and graphs will partially map back to the original, pre-XLA, graph.
- Tensorboard Improvements:
  - The Graph panel displays the correlation of an XLA cluster map to the original TensorFlow graph.
  - Reports now include Tensor shapes, dimensions, and data types.

Known Issues
- This software is only accessible in the NGC TensorFlow container.
- This software is only supported by TensorFlow 1.14
Chapter 7.
DEEP LEARNING PROFILER 19.09 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Key Features
The key features of DLProf v0.4.0/r19.09:

› CLI improvements: Report generation has changed, output options have been improved to make it easier for users to generate detailed reports.
› Tensorboard Improvements:
  › Model summary tab now shows a kernel summary report that details GPU time summary for kernels in use.
  › Iterations summary tab now shows operation names correlated with the kernels used in detail.

Known Issues
› This software is only accessible in the NGC TensorFlow container.
Chapter 8.
DEEP LEARNING PROFILER 19.08 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Key Features
The key features of DLProf v0.4.0/r19.08:

‣ Enabling faster generation of Tensorboard event files: Size of the protobuf file used for data collection is smaller so that more profiling data points can be collected and loading Tensorboard event files is faster.
‣ Kernel report showing usage per kernel: DLProf has added a new report showing CUDA Kernel usage for the benefit of advanced researchers trying to understand which kernels were run when model was trained.

Known Issues
‣ This software is only accessible in the NGC TensorFlow container.
Chapter 9.
DEEP LEARNING PROFILER 19.07 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Key Features
The key features of DLProf v0.3.9/r19.07:

- Ability to aggregate data per iteration: User can specify the iteration range to aggregate timing metrics for all reports by specifying start and stop iterations.
- Tensor Core Report: DLProf can create a CSV report listing all unique Tensor Core kernels that were executed in the model, along with node and timing metric information.
- Support for Tensorboard 1.14: Visualization component is now based on Tensorboard 1.14.

Known Issues

- This is early version software. It is only accessible in the NGC TensorFlow container.

Resolved Issues

- Bug fixes in timing data in Tensorboard.
- Fix for models not showing Tensor Core usage.
Chapter 10.
DEEP LEARNING PROFILER 19.06 RELEASE NOTES

Description
Deep Learning Profiler (DLProf) is a tool for profiling deep learning models to help data scientists understand and improve performance of their models visually via Tensorboard or by analyzing text reports. It also helps understand resource usage when models are trained.

Key Features
The key features of DLProf v0.3.7/r19.06:

‣ Easy Profiling: Just provide graphdef file and prefix training script with “dlprof” to automatically get Tensorboard event files with profile data.
‣ Visualization: Modified version of Tensorboard shows profiling data on a familiar interface.
‣ Reports: Several profiling reports can be generated to provide maximum value.

Known Issues
‣ This is a very early version software. It is only accessible in the NGC TensorFlow container. Currently we do not support profiling models with the XLA option. Some of the timing data in Tensorboard may not add up. The default key node of 'global_step/add' assumed by DLProf may not exist in some models. If no iterations are being detected, users should try running the Deep Learning Profiler with a different key node --key_node={new_key_node}.

Some models execute nodes in the final iterations that are not representative of a normal training iteration. These final iterations may not contain any GPU activity.
‣ DLProf and Nsight Systems in the container will not work with NVIDIA GPU Drivers newer than version 418.
Resolved Issues

There are no resolved issues in this release.
Notice

THE INFORMATION IN THIS GUIDE AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS GUIDE IS PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE INFORMATION FOR THE PRODUCT, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the product described in this guide shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

THE NVIDIA PRODUCT DESCRIBED IN THIS GUIDE IS NOT FAULT TOLERANT AND IS NOT DESIGNED, MANUFACTURED OR INTENDED FOR USE IN CONNECTION WITH THE DESIGN, CONSTRUCTION, MAINTENANCE, AND/OR OPERATION OF ANY SYSTEM WHERE THE USE OR A FAILURE OF SUCH SYSTEM COULD RESULT IN A SITUATION THAT THREATENS THE SAFETY OF HUMAN LIFE OR SEVERE PHYSICAL HARM OR PROPERTY DAMAGE (INCLUDING, FOR EXAMPLE, USE IN CONNECTION WITH ANY NUCLEAR, AVIONICS, LIFE SUPPORT OR OTHER LIFE CRITICAL APPLICATION). NVIDIA EXPRESSLY DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR SUCH HIGH RISK USES. NVIDIA SHALL NOT BE LIABLE TO CUSTOMER OR ANY THIRD PARTY, IN WHOLE OR IN PART, FOR ANY CLAIMS OR DAMAGES ARISING FROM SUCH HIGH RISK USES.

NVIDIA makes no representation or warranty that the product described in this guide will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this guide. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this guide, or (ii) customer product designs.

Other than the right for customer to use the information in this guide with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this guide. Reproduction of information in this guide is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

Trademarks

NVIDIA, the NVIDIA logo, and cuBLAS, CUDA, cuDNN, DALI, DIGITS, DGX, DGX-1, DGX-2, DGX Station, DLProf, Jetson, Kepler, Maxwell, NCCL, Nsight Compute, Nsight Systems, NvCaffe, PerfWorks, Pascal, SDK Manager, Tegra, TensorRT, TensorRT Inference Server, Tesla, TF-TRT, and Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2020 NVIDIA Corporation. All rights reserved.

www.nvidia.com