Deep Learning Profiler Release Notes

Release Notes

RN-09462-005_v21.12 | December 2021

Table of Contents

Chapter 1. Deep Learning Profiler v1.8.0 Release Notes

Description

DLProf Release for v1.8.0 is available in the 21.12 NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index. This will be the final release of DLProf. It will not be included in future NVIDIA NGC containers, but will still be available to download from the NVIDIA PY Index.

Driver Requirements

Release 21.12 is based on NVIDIA CUDA 11.5.0, which requires NVIDIA Driver release 495 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.8.0 are:

- Released in the TensorFlow 1.x 21.12, TensorFlow 2.x 21.12 and PyTorch 21.12 NGC containers.

Known Issues

- This software is only supported for TensorFlow 1.15.5, TensorFlow 2.6.2, PyTorch 1.11, and TensorRT 8.2.1.8.
- Collecting I/O data can cause issues in some cases, so it is off by default in this release. To turn it on, pass the following option to DLProf CLI:

```
--nsys_opts="-t cuda,nvtx,osrt -s cpu"
```

Resolved Issues

- None.

Chapter 2. Deep Learning Profiler 21.11 Release Notes

Description

DLProf Release for 21.11, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

DLProf v1.8, which will be included in the 21.12 DLFW containers, will be the final release of DLProf. Starting with the 22.01 containers, DLProf will no longer be included. It can still be manually installed via a pip wheel on the NVIDIA PY index.

Driver Requirements

Release 21.11 is based on NVIDIA CUDA 11.5.0, which requires NVIDIA Driver release 495 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.7.0 / r21.11 are:

- Released in the TensorFlow 1.x 21.11, TensorFlow 2.x 21.11 and PyTorch 21.11 NGC containers.
- Added support for new Group Op view in the DLProf Viewer.
- Removed support for custom NVTX domains.

Known Issues

- This software is only supported for TensorFlow 1.15., TensorFlow 2.6, PyTorch 1.10, and TensorRT 8.0.3.

- Collecting I/O data can cause issues in some cases, so it is off by default in this release. To turn it on, pass the following option to DLProf CLI:

```
--nsys_opts="-t cuda,nvtx,osrt -s cpu"
```

Resolved Issues

- None.

Chapter 3. Deep Learning Profiler 21.10 Release Notes

Description

DLProf Release for 21.10, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.10 is based on NVIDIA CUDA 11.4.2, which requires NVIDIA Driver release 470 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.6.0 / r21.10 are:

- Released in the TensorFlow 1.x 21.10, TensorFlow 2.x 21.10 and PyTorch 21.10 NGC containers.
- Added support for running DLProf on ARM ISA.
- Added I/O and Dataloader categories to Summary report and to Resource Usage Breakdown.

Known Issues

- This software is only supported for TensorFlow 1.15.5, TensorFlow 2.6.0, PyTorch 1.10, and TensorRT 8.0.3.
- Collecting I/O data can cause issues in some cases, so it is off by default in this release. To turn it on, pass the following option to DLProf CLI:

```
--nsys_opts="-t cuda,nvtx,osrt -s cpu"
```

Resolved Issues

- None.

Chapter 4. Deep Learning Profiler 21.09 Release Notes

Description

DLProf Release for 21.09, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.09 is based on NVIDIA CUDA 11.4.2, which requires NVIDIA Driver release 470 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.5.0 / r21.09 are:

- Released in the TensorFlow 1.x 21.09, TensorFlow 2.x 21.09 and PyTorch 21.09 NGC container.
- Added parameters and output shapes into the iteration report for GEMM and CONV operations for TensorFlow 1.x and 2.x.

Known Issues

- This software is only supported for TensorFlow 1.15.5, TensorFlow 2.6.0, PyTorch 1.10, and TensorRT 8.0.3.

Resolved Issues

- Fixed issue delaying writing many iterations to the DLProf database.

Chapter 5. Deep Learning Profiler 21.08 Release Notes

Description

DLProf Release for 21.08, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Note: NVTX-Plugins are being removed from the Deep Learning Frameworks TensorFLow 1.x & TensorFLow 2.x containers starting with the 21.09 release. They can still be manually installed from https://github.com/NVIDIA/nvtx-plugins.

Driver Requirements

- Requires DLProf SQLite database generated by DLProf v1.2 or later.
- Ensure that you have access and are logged into NGC. For step-by-step instructions, see the NGC Getting Started Guide.
- Install Docker and nvidia-docker. For DGX users, see Preparing to use NVIDIA Containers. For users other than DGX, see nvidia-docker installation documentation.

New Features

The key features of DLProf v1.4.0 / r21.08 are:

- Released in the TensorFlow 1.x 21.08, TensorFlow 2.x 21.08 and PyTorch 21.08 NGC container.
- Multi-GPU calculation and reporting enhancement.
- Added parameters and output shapes into the iteration report for GEMM and CONV operations.

Known Issues

- This software is only supported for TensorFlow 1.15, TensorFlow 2.34, PyTorch 1.9, and TensorRT 7.2.

- nvtx-plugins is being removed from the DLFW TensorFlow 1 & TensorFlow 2 containers starting with the 21.09 release.

Resolved Issues

- None.

Chapter 6. Deep Learning Profiler 21.07 Release Notes

Description

DLProf Release for 21.07, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.07 is based on NVIDIA CUDA 11.4.0, which requires NVIDIA Driver release 470 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.3.0 / r21.07 are:

- Released in the TensorFlow 1.x 21.07, TensorFlow 2.x 21.07 and PyTorch 21.07 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.5, TensorFlow 2.5.0, PyTorch 1.10.0, TensorRT 8.0, and Nsight Systems 2021.2.4.12.
- Added database versioning and backwards compatibility starting with v1.2 databases.
- Added underutilized GPU detector.

Known Issues

- This software is only supported for TensorFlow 1.15, TensorFlow 2.5.0, PyTorch 1.10.0, and TensorRT 8.0.

Resolved Issues

- Fixed issue where memory could get corrupted if not all GPUs were used

Chapter 7. Deep Learning Profiler 21.06 Release Notes

Description

DLProf Release for 21.06, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.06 is based on NVIDIA CUDA 11.3.1, which requires NVIDIA Driver release 465.19.01 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.40 (or later R418), 440.33 (or later R440), 450.51 (or later R450), or 460.27 (or later R460). 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 v1.2.0 / r21.06 are:

- Latest DLProf build is based on TensorFlow 1.15.5, TensorFlow 2.4.0, PyTorch 1.9.0, TensorRT 7.2.3, and Nsight Systems 2021.2.1.58.
- Resource Usage Breakdown: split profile activity into resource categories.
- Aggregation performance improvements.

Known Issues

- This software is only supported for TensorFlow 1.15, TensorFlow 2.4, PyTorch 1.9, and TensorRT 7.2.

Resolved Issues

- None

Chapter 8. Deep Learning Profiler 21.04 Release Notes

Description

DLProf Release for 21.04, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.04 is based on NVIDIA CUDA 11.3.0, which requires NVIDIA Driver release 465.19.01 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.xx, 450.xx, or 455.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 v1.1.0 / r21.04 are:

- Released in the TensorFlow 1.x 21.04, TensorFlow 2.x 21.04 and PyTorch 21.04 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.5, TensorBoard 1.15.0, TensorFlow 2.3.1, TensorBoard 2.3.0, PyTorch 1.8.0, and Nsight Systems 2020.4.3.
- Improved DLProf database
  - Redundant data removed to reduce size
  - Improve write speed
  - Support for multiple aggregations in the same database

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.

- This software is only supported for TensorFlow 1.15, TensorFlow 2.3, PyTorch 1.8, TensorBoard 1.15, and TensorBoard 2.3.
- When launching TensorBoard in a TensorFlow 2.x container, the --bind_all argument must be passed onto the command line. Example:

```
# tensorboard --bind_all --logdir /path/to/event_files
```

Resolved Issues

- None

Chapter 9. Deep Learning Profiler 21.03 Release Notes

Description

DLProf Release for 21.03, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.03 is based on NVIDIA CUDA 11.2.1, which requires NVIDIA Driver release 460.32.03 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.30, 450.xx, or 455.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 v1.0.0 / r21.03 are:

- Released in the TensorFlow 1.x 21.03, TensorFlow 2.x 21.03 and PyTorch 21.03 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.5, TensorBoard 1.15.0, TensorFlow 2.3.1, TensorBoard 2.3.0, PyTorch 1.8.0, and Nsight Systems 2020.4.3.
- Expert Systems can detect when GPU memory is underutilized and recommend increasing batch size.
- DLProf can now profile TensorRT models.
- DLProf can detect NCCL events and properly associate GPU activity to them.

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.

- This software is only supported for TensorFlow 1.15, TensorFlow 2.3, PyTorch 1.8, TensorBoard 1.15, and TensorBoard 2.3.
- When profiling on multi-gpu, in a rare case DLProf can get stuck after printing, `DLprof completed system call successfully`. The recommended workaround is to run on a single GPU system or use version 21.02.
- When launching TensorBoard in a TensorFlow 2.x container, the --bind_all argument must be passed onto the command line. Example:

```
# tensorboard --bind_all --logdir /path/to/event_files
```

Resolved Issues

- None

Chapter 10. Deep Learning Profiler 21.02 Release Notes

Description

DLProf release for 21.02, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 21.02 is based on NVIDIA CUDA 11.2, which requires NVIDIA Driver release 455 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.30, or 450.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 v0.19.0 / r21.02 are:

- Released in the TensorFlow 1.x 21.02, TensorFlow 2.x 21.02 and PyTorch 21.02 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.4, TensorBoard 1.15.0, TensorFlow 2.3.1, TensorBoard 2.3.0, PyTorch 1.8.0, and Nsight Systems 2020.3.4.
- Added support for DALI.
- Improved PyTorch profiling
  - Iteration reports can print the long kernel names
  - Iteration reports display IO bytes
  - Expert Systems detector will detect when emit_nvtx is not used
- Capture all unassociated CUDA operations in all frameworks.

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.
- This software is only supported for TensorFlow 1.15, TensorFlow 2.3, PyTorch 1.8, TensorBoard 1.15, and TensorBoard 2.3.
- When launching TensorBoard in a TensorFlow 2.x container, the --bind_all argument must be passed onto the command line. Example:

```
# tensorboard --bind_all --logdir /path/to/event_files
```

Resolved Issues

- None

Chapter 11. Deep Learning Profiler 21.01 Release Notes

Description

The NVIDIA release for Deep Learning Profiler 21.01 has been canceled. The next release will be the 21.02 release which is expected to be released at the end of February.

Chapter 12. Deep Learning Profiler 20.12 Release Notes

Description

DLProf release for 20.12, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 20.12 is based on NVIDIA CUDA 11.1.1, which requires NVIDIA Driver release 455 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.30, or 450.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 v0.18.0 / r20.12 are:

- Released in the TensorFlow 1.x 20.12, TensorFlow 2.x 20.11 and PyTorch 20.11 NGC container
- Latest DLProf build is based on TensorFlow 1.15.4, TensorBoard 1.15.0, TensorFlow 2.3.1, TensorBoard 2.3.0PyTorch 1.8.0, and Nsight Systems 2020.3.4.
- Multi-GPU profiling and reporting is now fully supported.
- Improved identification and association for each operation in reports with a unique operation ID.

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.
- This software is only supported for TensorFlow 1.15, TensorFlow 2.3, PyTorch 1.8, TensorBoard 1.15, and TensorBoard 2.3.

Resolved Issues

- PyTorch AMP operations that use Tensor Cores now report the correct datatype.

Chapter 13. Deep Learning Profiler 20.11 Release Notes

Description

DLProf release for 20.11, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers, and as a Python Wheel on the NVIDIA PY Index.

Driver Requirements

Release 20.11 is based on NVIDIA CUDA 11.1.0, which requires NVIDIA Driver release 455 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.30, or 450.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 v0.17.0 / r20.11 are:

- Released in the TensorFlow 1.x 20.11, TensorFlow 2.x 20.11 and PyTorch 20.11 NGC container
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.7.0, and Nsight Systems 2020.3.2.
- Support for profiling of models using TensorFlow 2.x, starting with TensorFlow 2.3.1.

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.
- This software is only supported for TensorFlow 1.15, TensorFlow 2.3, PyTorch 1.7, TensorBoard 1.15, and TensorBoard 2.3.

Resolved Issues

- Expert Systems correctly detects AMP usage for PyTorch models.

Chapter 14. Deep Learning Profiler 20.10 Release Notes

Description

DLProf release for 20.10, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers.

Driver Requirements

Release 20.10 is based on NVIDIA CUDA 11.1.0, which requires NVIDIA Driver release 455 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 418.xx, 440.30, or 450.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 v0.16.0 / r20.10 are:

- Released in the TensorFlow 1.x 20.10, TensorFlow 2.x 20.10 and PyTorch 20.10 NGC container
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.6.0, and Nsight Systems 2020.3.2.
- Tensor shape constraints recommendations are not being reported because they no longer prohibit tensor core usage.
- When using XLA mode, a generic AMP usage recommendation will be displayed.

Known Issues

- This software is accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel.
- This software is only supported for TensorFlow 1.15, and PyTorch 1.6 and TensorBoard 1.15.

- Simple mode only profiling is supported for TensorFlow 2.
- The PyTorch AMP datatype is not correctly detected and causes the Expert Systems to recommend AMP in PyTorch even when AMP is enabled.

Resolved Issues

- None

Chapter 15. Deep Learning Profiler 20.09 Release Notes

Description

DLProf release for 20.09, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers.

Driver Requirements

Release 20.09 is based on NVIDIA CUDA 11.0.3, which requires NVIDIA Driver release 450 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 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 v0.15.0 / r20.09 are:

- Released in the TensorFlow 1.x 20.09, TensorFlow 2.x 20.09 and PyTorch 20.09 NGC container
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.6.0, and Nsight Systems 2020.3.2.
- A stand-alone version of DLProf is available as a PIP wheel file.
- Expert Systems is now able to auto-detect and recommend an ideal iterations range
- PyTorch features:
  - Expert Systems detects the use of slow debugging API calls and recommends that these calls not be used.
  - DLProf recognizes the profiler.start and profiler.stop parameters in PyTorch scripts to set specific profile ranges.

Known Issues

- This software is only accessible in the NGC TensorFlow and PyTorch containers and as a separate PIP wheel in the nvidia-pyindex.
- This software is only supported for TensorFlow 1.15, and PyTorch 1.6 and TensorBoard 1.15.
- Partial simple mode profiling is supported for TensorFlow 2.

Resolved Issues

- None

Chapter 16. Deep Learning Profiler 20.08 Release Notes

Description

DLProf release for 20.08, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers.

Driver Requirements

Release 20.08 is based on NVIDIA CUDA 11.0.194, which requires NVIDIA Driver release 450 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 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 v0.14.0 / r20.08 are:

- Released in the TensorFlow 1.x 20.08, TensorFlow 2.x 20.08and PyTorch 20.08 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.6.0, and Nsight Systems 2020.3.2.
- DLProf support for PyTorch is now feature equivalent to the DLProf support for TensorFlow 1.x:
  - Correctly identifies forward and backwards operations.
  - Tensor Core eligible operations list corrected for improved accuracy.
  - Stack trace and direction added to reports.
  - Data loader detector improved to provide more details.
  - Methodology added to differentiate between calls to the same function from the same stack trace.

- Testing and detection of Tensor Core kernels improved and provides a high level of confidence in TC usage metrics across all frameworks.
- Tensor Core utilization calculation improved for XLA runs.

Known Issues

- This software is only accessible in the NGC TensorFlow and PyTorch containers.
- This software is only supported for TensorFlow 1.15, and PyTorch 1.6 and TensorBoard 1.15.
- Partial simple mode profiling is supported for TensorFlow 2.

Resolved Issues

- None

Chapter 17. Deep Learning Profiler 20.07 Release Notes

Description

DLProf release for 20.07, available in the NVIDIA TensorFlow 1.x, TensorFlow 2.x, and PyTorch NGC containers.

Driver Requirements

Release 20.07 is based on NVIDIA CUDA 11.0.167, which requires NVIDIA Driver release 450 or later. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 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 v0.13.0 / r20.07 are:

- Released in the TensorFlow 1.x 20.07, TensorFlow 2.x 20.07 and PyTorch 20.07 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.6.0, and Nsight Systems 2020.3.2.
- Support for Tensor core detection in the new A100 GPU architecture, new TF32 tensorcores are detected for Tensorflow and Pytorch model profiles.
- Improved PyTorch support:
  - Detection of data loading bottlenecks by Expert systems.
  - Detection of AMP usage by Expert Systems.
  - Improved TensorCore utilization calculation.
- Tensorflow AMP enablement detection and recommendation in Expert Systems.
- Removed all dependencies on additional dynamic libraries.
- Added ability to specify name of model profiled:

- Model name is added to reports.
- Model name is propagated to the TensorBoard DLProf plugin.

Known Issues

- This software is only accessible in the NGC TensorFlow and PyTorch containers.
- This software is only supported for TensorFlow 1.15, and PyTorch 1.6 and TensorBoard 1.15.
- Partial simple mode profiling is supported for TensorFlow 2.

Resolved Issues

- None

Chapter 18. Deep Learning Profiler 20.06 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.

Driver Requirements

Release 20.06 is based on NVIDIA CUDA 11.0.167, which requires NVIDIA Driver release 450.36.06. However, if you are running on Tesla (for example, T4 or any other Tesla board), you may use NVIDIA driver release 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 v0.12.0 / r20.06 are:

- Released in the TensorFlow 20.06 and PyTorch 20.06 NGC container.
- Latest DLProf build is based on TensorFlow 1.15.2, TensorBoard 1.15.0, PyTorch 1.6.0, and Nsight Systems 2020.1.1.
- PyTorch support for DLProf build released in the PyTorch NGC container.
  - Depends on of PyProf and Nsight Systems which are provided in the container.
  - Generates all reports and a GPU Event file that can be viewed with the NVIDIA DLProf Tensorboard Plugin is included for Tensorboard visualization.
- New command line switch options.
  - Align CLI switch options and style with Nsight Systems CLI.
  - Improved help messages from `dlprof --help`.

  - Known issue: boolean options now require specifying `true` or `false`, e.g. `--force=true`.
- Support for multiple Frameworks and NGC containers.
  - Each supported NGC container has a customized DLProf that is optimized for the Deep Learning framework in the container.
  - Supported Frameworks:
    - Tensorflow 1.x
    - Tensorflow 2.x - simple mode only
    - PyTorch
- Recommendations for shape and data type are improved in Tensorboard and on command line.
- Reporting enhancements
  - Print all reports with `--reports=all`
  - Added start and stop iterations to Summary report.
  - Added more detailed TensorCore GPU time to the Tensor report.
  - Fixed issue in kernel report that was over-counting the number of unique kernels.

Known Issues

- This software is only accessible in the NGC TensorFlow and PyTorch containers.
- This software is only supported for TensorFlow 1.15, and PyTorch 1.6 and TensorBoard 1.15.

Resolved Issues

- None

Chapter 19. 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.

Chapter 20. Deep Learning Profiler 20.02 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 20.02 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 21. 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 22. 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 23. 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 24. 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 25. 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 26. 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 27. 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 28. 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

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform 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 document. NVIDIA accepts no 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 document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
https://www.nvidia.com

Trademarks

NVIDIA, the NVIDIA logo, and cuBLAS, CUDA, DALI, DGX, DGX-1, DGX-2, DGX Station, DLProf, Jetson, Kepler, Maxwell, NCCL, Nsight Compute, Nsight Systems, NvCaffe, PerfWorks, Pascal, SDK Manager, Tegra, TensorRT, Triton 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

© 2021-2024 NVIDIA Corporation & Affiliates. All rights reserved.

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
https://www.nvidia.com