
DXT Explorer

Release 1.2

Jean Luca Bez, Suren Byna

Apr 14, 2022

GETTING STARTED

1	Build Instructions	3
1.1	Dependencies	3
1.2	Docker Image	3
2	Exploring	5
3	Ways to contribute	7
3.1	Reporting bugs	7
3.2	Suggesting enhancements	7
3.3	Testing	7
4	Citation	9
5	Copyright	11
6	License Agreement	13

DXT Explorer is an interactive web-based log analysis tool to visualize Darshan DXT logs and help understand the I/O behavior of applications. Our tool adds an interactive component to Darshan trace analysis that can aid researchers, developers, and end-users to visually inspect their applications' I/O behavior, zoom-in on areas of interest and have a clear picture of where is the I/O problem.

BUILD INSTRUCTIONS

1.1 Dependencies

DXT Explorer requires a Darshan log file collected with tracing data. The Darshan eXtended Tracing (DXT) support is disabled by default in Darshan. To enable tracing globally for all files, you need to set the `DXT_ENABLE_IO_TRACE` environment variable as follows:

```
export DXT_ENABLE_IO_TRACE=1
```

To enable tracing for particular files you can refer to the Darshan's documentation page.

To use DXT Explorer, you need to have Python 3 and R already installed in your system, and install some required Python libraries:

```
pip install -r requirements.txt
```

In the first execution ever, DXT Explorer will automatically download any missing R packages required, thus it might take longer to generate the plot. This is all done at user level, without any need for elevated privileges.

You also need to have Darshan Utils installed (`darshan-dxt-parser`) and available in your path.

Note: In Summit, if you want to run DXT Explorer, you need to load some modules:

```
module load python r cairo
```

1.2 Docker Image

You can also use a Docker image already pre-configured with all dependencies to run DXT Explorer:

```
docker pull hpcio/dxt-explorer
```

Since we need to provide an input file and access the generated `.html` files, make sure you are mounting your current directory in the container and removing the container after using it. You can pass the same arguments described above, after the container name (`dxt-explorer`).

```
docker run --rm --mount \
  type=bind,source="$(PWD)",target="/dxt-explorer/darshan" \
  dxt-explorer darshan/<FILE>.darshan
```


EXPLORING

Once you have the dependencies and DXT Explorer installed, you can run:

```
dxt-explore DARSHAN_FILE_COLLECTED_WITH_DXT_ENABLE.darshan
```

```
usage: dxt-explorer [-h] [-o OUTPUT] [-t] [-s] [-d] [-l] [--start START] [--end END] [--from START_RANK] [--to END_RANK] [--browser] darshan
```

DXT Explorer:

positional arguments:

darshan Input .darshan file

optional arguments:

-h, --help show this help message and exit
-o OUTPUT, --output OUTPUT Name of the output file
-t, --transfer Generate an interactive data transfer explorer
-s, --spatiality Generate an interactive spatiality explorer
-d, --debug Enable debug mode
-l, --list List all the files with trace
--start START Report starts from X seconds (e.g., 3.7) from beginning of the job
--end END Report ends at X seconds (e.g., 3.9) from beginning of the job
--from START_RANK Report start from rank N
--to END_RANK Report up to rank M
--browser Open the browser with the generated plot

DXT Explorer will generate by default a `explore.html` file with an interactive plot that you can open in any browser to explore. If you enabled the transfer or spatiality plots, additional `.html` files will be generated, one for each type. You are expected to visualize the following messages in the console:

```
2021-10-05 03:21:34,907 explore - INFO - darshan-dxt-parser: FOUND
2021-10-05 03:21:34,907 explore - INFO - Rscript: FOUND
2021-10-05 03:21:34,907 explore - INFO - parsing darshan/<FILE>.darshan file
2021-10-05 03:21:35,248 explore - INFO - generating an intermediate CSV file
2021-10-05 03:21:36,240 explore - INFO - generating interactive operation plot
2021-10-05 03:21:54,657 explore - INFO - SUCCESS
```

You can find a couple of interactive examples of DXT traces collected from FLASH, E2E, and OpenPMD in the [companion repository](#) for our PDSW'21 paper.

Note: If you do not want to install DXT Explorer but rather prefer to run it as a standalone script, you can use the following command to launch it. However, remember that you are still required to install all dependencies.

```
python3 explore/dxt.py YOUR-DXT-TRACE.darshan
```

WAYS TO CONTRIBUTE

We appreciate your interest in **DXT Explorer**, and thank you for taking the time to contribute!

We have compiled a set of instructions to help us make DXT Explorer even better.

3.1 Reporting bugs

You can open a new issue using our GitHub [issue tracker](#). If you run into an issue, please search first to ensure the issue has not been reported before. Open a new issue only if you have not found anything similar to your issue. Please, try to provide as much information as possible to reproduce your bug quickly.

3.2 Suggesting enhancements

You can use our GitHub [issue tracker](#) to describe your proposed feature. Please, provide the necessary context, covering why it is needed and what problem does it solve.

3.3 Testing

DXT Explorer constantly receives updates and improvements. If you can run the latest version, please consider helping us by reporting your findings, including bugs and performance regressions. Running DXT Explorer with different configurations and platforms helps us a lot in making it more robust by quickly identifying and solving issues.

CITATION

You can find more information about DXT Explorer in our PDSW'21 paper. If you use DXT in your experiments, please consider citing:

```
@inproceedings{dxt-explorer,  
  title = {{I/O Bottleneck Detection and Tuning: Connecting the Dots using Interactive_  
↵Log Analysis}},  
  author = {Bez, Jean Luca and Tang, Houjun and Xie, Bing, and Williams-Young, David_  
↵and Latham, Rob and Ross, Rob and Oral, Sarp and Byna, Suren},  
  booktitle = {2021 IEEE/ACM 6th International Parallel Data Systems Workshop (PDSW)}  
  year = {2021}  
}
```


COPYRIGHT

DXT Explorer Copyright (c) 2021, The Regents of the University of California, through Lawrence Berkeley National Laboratory (subject to receipt of any required approvals from the U.S. Dept. of Energy). All rights reserved.

If you have questions about your rights to use or distribute this software, please contact Berkeley Lab's Intellectual Property Office at IPO@lbl.gov.

Attention: This Software was developed under funding from the U.S. Department of Energy and the U.S. Government consequently retains certain rights. As such, the U.S. Government has been granted for itself and others acting on its behalf a paid-up, nonexclusive, irrevocable, worldwide license in the Software to reproduce, distribute copies to the public, prepare derivative works, and perform publicly and display publicly, and to permit others to do so.

LICENSE AGREEMENT

DXT Explorer Copyright (c) 2021, The Regents of the University of California, through Lawrence Berkeley National Laboratory (subject to receipt of any required approvals from the U.S. Dept. of Energy). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- (1) Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- (2) Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- (3) Neither the name of the University of California, Lawrence Berkeley National Laboratory, U.S. Dept. of Energy nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

You are under no obligation whatsoever to provide any bug fixes, patches, or upgrades to the features, functionality or performance of the source code (“Enhancements”) to anyone; however, if you choose to make your Enhancements available either publicly, or directly to Lawrence Berkeley National Laboratory, without imposing a separate written license agreement for such Enhancements, then you hereby grant the following license: a non-exclusive, royalty-free perpetual license to install, use, modify, prepare derivative works, incorporate into other computer software, distribute, and sublicense such enhancements or derivative works thereof, in binary and source code form.